



19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA

11 Número de publicación: **2 340 154**

51 Int. Cl.:

C12N 15/12 (2006.01)

C12N 15/63 (2006.01)

C07K 14/435 (2006.01)

C12N 15/62 (2006.01)

C07K 16/18 (2006.01)

A01K 67/00 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Número de solicitud europea: **03706812 .9**

96 Fecha de presentación : **17.01.2003**

97 Número de publicación de la solicitud: **1485481**

97 Fecha de publicación de la solicitud: **15.12.2004**

54

Título: **Proteínas fluorescentes de *Aequorea coerulescens* y uso de las mismas.**

30

Prioridad: **22.01.2002 US 351518 P**

73

Titular/es: **EVROGEN IP
16/10 Miklukho-Maklaya Street
Moscow 117997, RU**

45

Fecha de publicación de la mención BOPI:
31.05.2010

72

Inventor/es: **Gurskaya, Nadejda, G.;
Fradlov, Arkadiy, F.;
Lukyanov, Sergey, A. y
Punkova, Natalia**

45

Fecha de la publicación del folleto de la patente:
31.05.2010

74

Agente: **Ungría López, Javier**

ES 2 340 154 T3

Aviso: En el plazo de nueve meses a contar desde la fecha de publicación en el Boletín europeo de patentes, de la mención de concesión de la patente europea, cualquier persona podrá oponerse ante la Oficina Europea de Patentes a la patente concedida. La oposición deberá formularse por escrito y estar motivada; sólo se considerará como formulada una vez que se haya realizado el pago de la tasa de oposición (art. 99.1 del Convenio sobre concesión de Patentes Europeas).

DESCRIPCIÓN

Proteínas fluorescentes de *Aequorea coerulescens* y uso de las mismas.

5 Campo de la invención

Esta invención se refiere a proteínas fluorescentes.

Antecedentes de la invención

10 El etiquetado es una herramienta para el mareaje de una proteína, célula, u organismo de interés y desempeña un papel destacado en muchas aplicaciones diagnósticas bioquímicas, de biología molecular y médicas. Se han desarrollado y utilizado en la técnica varias etiquetas diferentes, incluyendo etiquetas radiactivas, cromoetiquetas, etiquetas fluorescentes, etiquetas quimioluminiscentes, y similares, con diferentes propiedades y usos óptimos. No obstante, 15 continúa el interés por el desarrollo de nuevas etiquetas. De particular interés es el desarrollo de nuevas etiquetas de proteína, incluyendo etiquetas de proteínas fluorescentes.

Literatura relevante

20 Las Patentes de los Estados Unidos de interés incluyen: 6.066.476; 6.020.192; 5.985.577; 5.976.796; 5.968.750; 5.968.738; 5.958.713; 5.919.445; 5.874.304 y 5.491.084. Las Publicaciones de Patentes Internacionales de interés incluyen: WO 00/46233; WO 99/49019; y DE 197 18 640 A. También tienen interés: Anderluh *et al.*, Biochemical and Biophysical Research Communications (1996) 220:437-442; Dove *et al.*, Biological Bulletin (1995) 189:288-297; Fradkov *et al.*, FEBS Lett. (2000) 479(3):127-30; Gurskaya *et al.*, FEBS Lett., (2001) 507(1):16-20; Gurskaya *et al.*, 25 BMC Biochem. (2001) 2:6; Lukyanov, K., *et al.* (2000) J. Biol. Chemistry 275(34):25879-25882; Macek *et al.*, Eur. J. Biochem. (1995) 234:329-335; Martynov *et al.*, J. Biol. Chem. (2001) 276:21012-6; Matz, M.V., *et al.* (1999) Nature Biotechnol., 17:969-973; Tersikh *et al.*, Science (2000) 290:1585-8; Tsien, Annual Rev. of Biochemistry (1998) 67:509-544; Tsien, Nat. Biotech. (1999) 17:956-957; Ward *et al.*, J. Biol. CHEM. (1979) 254:781-788; Wiedermann *et al.*, Jahrestagung der Deutschen Gesellschaft für Tropenökologie-gto Ulm. 17-19.02.1999. Poster P-4.20; Yarbrough 30 *et al.*, Proceedings of the National Academy of Sciences (2001) 98:462-7

Prasher *et al.*, (1992), Gene, Vol 111 págs. 229-233, describen la estructura primaria de la proteína fluorescente verde de *Aequorea victoria* incluyendo sus secuencias de ADN y de péptidos.

35 Inouye y Tsuji (1999) FEBS Letters, Vol. 341 págs. 277-280, describen GFP de *Aequorea victoria*, y caracterizan las características de fluorescencia de las proteínas recombinantes.

Compendio de la invención

40 La presente invención proporciona composiciones de ácido nucleico que codifican una proteína incolora única de *Aequorea coerulescens* y sus mutantes fluorescentes, así como las proteínas y péptidos codificados por los ácidos nucleicos. Las proteínas de la presente invención son proteínas que están coloreadas y/o son fluorescentes y/o pueden ser fotoactivadas, donde esta característica óptica surge de la interacción de dos o más residuos aminoácido de la proteína. También tienen interés las proteínas que son proteínas que tienen al menos 95% de identidad con las proteínas específicas referidas anteriormente incluyendo las proteínas de fusión que incorporan péptidos de la presente invención, así 45 como los anticuerpos para estas proteínas. La proteína sujeto y las composiciones de ácidos nucleicos encuentran uso en una variedad de aplicaciones diferentes.

Descripción de las figuras

50 La figura 1 es la secuencia de aminoácidos y la secuencia de ácido nucleico que codifica la proteína GFP de tipo salvaje de *Aequorea coerulescens*, referida en la presente memoria como acGFP.

55 La Figura 2 es la comparación de las secuencias de aminoácidos de GFP de *Aequorea victoria* y acGFP de *Aequorea Coerulescens*.

La Figura 3 es la secuencia de aminoácidos y la secuencia de ácido nucleico que codifican la acGFP mutante, Z1.

La Figura 4 son los espectros de excitación-emisión de Z1 mutante.

60 La Figura 5 es la secuencia de aminoácidos y la secuencia de ácido nucleico de la secuencia de acGFP mutante, Z2.

La Figura 6 es la secuencia de aminoácidos y la secuencia de ácido nucleico que codifican la acGFP mutante, G1.

65 La Figura 7 es la secuencia de aminoácidos y de ácido nucleico que codifican la acGFP mutante, G2.

La Figura 8 son los espectros de excitación-emisión de G2 mutante.

ES 2 340 154 T3

La Figura 9 es la secuencia de aminoácidos y la secuencia de ácido nucleico que codifican la acGFP mutante, G22.

La Figura 10 son los espectros de excitación-emisión de G22 mutante.

5 La Figura 11 análisis de electroforesis en gel de las proteínas acGFP de tipo salvaje y acGFP mutante.

La Figura 12 es la secuencia de aminoácidos y la secuencia de ácido nucleico que codifican la acGFP mutante, G22-G222E.

10 La Figura 13 son los espectros de absorción excitación-emisión del mutante G22-G222E.

La Figura 14 proporciona los espectros de fotoconversión inducida por UV de G22-G222E.

15 La Figura 15 es la secuencia de aminoácidos y la secuencia de ácido nucleico que codifican la acGFP mutante, G22-G222E/Y22 0L.

La Figura 16 son los espectros de excitación-emisión de G22-G222E/Y220L mutante.

20 La Figura 17 es la secuencia de aminoácidos y la secuencia de ácido nucleico que codifican la acGFP mutante, 220-II-5.

La Figura 18 proporciona las propiedades espectrales de 220-II-5 mutante.

25 La Figura 19 es la secuencia de aminoácidos y la secuencia de ácido nucleico que codifican la acGFP mutante, CFP-rand3.

La Figura 20 son los espectros de excitación-emisión de CFP-rand3 mutante.

30 La Figura 21 es la secuencia de aminoácidos y la secuencia de ácido nucleico que codifican acGFP mutante, CFP-3.

La Figura 22 proporciona las propiedades espectrales de CFP-3 mutante.

35 La Figura 23 son las secuencias de aminoácidos y de ácido nucleico de una versión humanizada de G22 mutante.

La Figura 24 muestra microfotografías de células de mamíferos que expresan G22 mutante.

La Figura 25 muestra la fotoactivación de 220-11-5 mutante en colonias de *E. coli*.

40 Descripción de la invención

La invención sujeto proporciona un ácido nucleico, donde el ácido nucleico codifica una proteína fluorescente, acGFP. En algunas realizaciones, el ácido nucleico es aislado, ha sido diseñado o está presente en un entorno distinto de su entorno natural. En algunas realizaciones, el ácido nucleico tiene una secuencia de residuos que comprende
45 las que se muestran en las figuras 01, 03, 05, 07, 09, 11, 13, 15, 17, 19, 21, o 23. En algunas realizaciones, el ácido nucleico de la presente invención codifica una proteína que tiene una secuencia de aminoácidos seleccionada del grupo que consiste en las que se muestran en las figuras 02, 04, 06, 08, 10, 12, 14, 16, 18, 20, 22.

También se proporcionan los constructos que comprenden un vector y un ácido nucleico de la presente invención.
50 Además, la presente invención proporciona casetes de expresión que incluyen: una región de iniciación de la transcripción funcional en un anfitrión de expresión, un ácido nucleico de la presente invención, y una región de terminación de la transcripción funcional en la casete de expresión como parte de un elemento extracromosómico o integrado en el genoma de la célula como resultado de la introducción de dicha casete de expresión en la célula.

55 También se proporcionan métodos de producción de una proteína cromogénica y/o fluorescente que incluyen el crecimiento de una célula de la presente invención, que expresa la proteína en la célula, y el aislamiento de la proteína esencialmente libre de otras proteínas.

60 Además, se proporcionan las proteínas codificadas por un ácido nucleico de la presente invención, puesto que son anticuerpos que se unen específicamente a las proteínas o péptidos de la presente invención.

Además, se proporcionan las células transgénicas (o su progenie) que incluyen un ácido nucleico de la presente invención, ya que son organismos transgénicos no humanos que incluyen un ácido nucleico de la presente invención.

65 También se proporcionan métodos que emplean una cromoproteína o una proteína fluorescente de la presente invención, o que emplean a un ácido nucleico que codifica una proteína cromogénica o fluorescente de la presente invención.

De acuerdo con la presente invención se pueden emplear técnicas de biología molecular, microbiología y de ADN recombinante convencionales en el conocimiento práctico de la técnica. Tales técnicas se explican con detalle en las publicaciones especializadas. Véanse, por ejemplo, Maniatis, Fritsch & Sambrook, *Molecular Cloning: A Laboratory Manual* (1982); *DNA Cloning: A Practical Approach*, Volumes I and II (D.N. Glover, ed. 1985); *Oligonucleotide Synthesis* (M.J. Gait, ed. 1984); *Nucleic Acid Hybridization* (B.D. Hames & S.J. Higgins, eds. (1984)); *Animal Cell Culture* (R.I. Freshney, ed. (1986)); *Immobilized Cells and Enzymes* (IRL Press (1986)); and B. Perbal, *A Practical Guide to Molecular Cloning* (1984).

Un “vector” es un replicón, tal como un plásmido, fago o cósmido, al que se puede conectar otro segmento de ADN.

Una “molécula de ADN” hace referencia a la forma polimérica de los desoxirribonucleótidos (adenina, guanina, timina o citosina) ya sea en forma de una sola cadena o de hélice bicatenaria. Este término se refiere sólo a la estructura primaria y secundaria de la molécula, y no se limita ninguna forma terciaria concreta. Así, “molécula de ADN” incluye ADN bicatenario encontrado, entre otros, en moléculas de ADN lineal (por ejemplo, fragmentos de restricción), virus, plásmidos y cromosomas.

Una “secuencia codificante” de ADN es una secuencia de ADN que es transcrita y traducida a un polipéptido *in vivo* cuando se coloca bajo el control de las secuencias reguladoras apropiadas. Los límites de la secuencia codificante son determinados por un codón de iniciación en el extremo 5’ (amino) y un codón de terminación de la traducción en el extremo 3’ (carboxilo). Una secuencia codificante puede incluir, pero no está limitada a, secuencias procarióticas, ADNc de ARNm eucariótico, secuencias de ADN genómico de ADN eucariótico (por ejemplo, mamíferos), y secuencias de ADN sintético. Una señal de poliadenilación y la secuencia de terminación de la transcripción se pueden localizar 3’ con respecto a la secuencia codificante.

Según se utiliza en la presente memoria, el término “hibridación” se refiere al proceso de asociación de dos cadenas de ácidos nucleicos para formar un dúplex antiparalelo estabilizado por medio de enlaces de hidrógeno entre los residuos de cadenas opuestas de ácido nucleico.

El término “oligonucleótido” se refiere a una molécula corta de ácido nucleico (menos de 100 bases de longitud).

Según se utiliza en la presente memoria, “secuencias reguladoras de ADN” son secuencias de control de la transcripción y la traducción, tales como promotores, potenciadores, señales de poliadenilación, terminadores, y similares, que proporcionan y/o regulan la expresión de una secuencia codificante en una célula anfitriona.

Una “secuencia promotora” es una región reguladora de ADN capaz de unirse a la ARN polimerasa en una célula e iniciar la transcripción de una secuencia codificante. Por ejemplo, la secuencia promotora se puede unir a su extremo 3’ por el sitio de iniciación de la transcripción y se extiende aguas arriba (5’) para incluir el número mínimo de bases o elementos necesarios para iniciar la transcripción a niveles detectables por encima del fondo. Dentro de la secuencia promotora se puede encontrar un sitio de iniciación de la transcripción, así como dominios de unión a proteínas responsables de la unión de la ARN polimerasa. Los promotores eucarióticos con frecuencia, aunque no siempre, contendrán cajas “TATA” y cajas “CAT”. Varios promotores, incluidos los promotores inducibles, se puede utilizar para conducir los diversos vectores de la presente invención.

Según se utilizan en la presente memoria, los términos “endonucleasas de restricción” y “enzimas de restricción” se refieren a las enzimas bacterianas, cada una de las cuales cortan el ADN bicatenario en o cerca de una secuencia de nucleótidos específica.

Una célula ha sido “transformada” o “transfectada” por el ADN exógeno o heterólogo cuando tal ADN ha sido introducido dentro de la célula. El ADN transformante puede ser o puede no ser integrado (unido covalentemente) en el genoma de la célula. En las células procariotas, de levaduras y de mamíferos, por ejemplo, el ADN transformante se puede mantener en un elemento episómico tal como un plásmido. Con respecto a las células eucarióticas, una célula transformada de forma estable es aquella en la que el ADN transformante se ha integrado en un cromosoma de la célula anfitriona o se mantiene extracromosómicamente de manera que el ADN transformante es heredado por las células hijas durante la replicación celular. Dicha célula eucariota transformada establemente es capaz de establecer líneas celulares o clones compuestos por una población de células hijas que contienen el ADN transformante. Un “clon” es una población de células derivadas de una sola célula o ancestro común por mitosis. Una “línea celular” es un clon o una célula que es capaz de un crecimiento estable *in vitro* durante muchas generaciones.

Una región “heteróloga” de un constructo de ADN es un segmento identificable de ADN dentro de una molécula de ADN más grande que no se encuentra asociado con la molécula más grande en la naturaleza; por ejemplo, cuando la región heteróloga codifica un ADN genómico de mamífero en el genoma de un organismo no mamífero. En otro ejemplo, el ADN heterólogo incluye secuencias codificantes en un constructo donde las porciones de genes de dos fuentes diferentes se han reunido para producir un producto proteico de fusión.

Según se utiliza en la presente memoria, el término “gen informador” se refiere a una secuencia codificante anclada al promotor heterólogo o a los elementos intensificadores y cuyo producto se puede analizar fácilmente y cuantificablemente cuando el constructo es introducido en los tejidos o las células.

Se prefiere que los aminoácidos descritos en la presente memoria estén en la forma isomérica “L”. Las secuencias de aminoácidos se dan en un código de una letra (A: alanina; C: cisteína, D: ácido aspártico, E: ácido glutámico, F: fenilalanina, G: glicina, H: histidina, I: isoleucina, K: lisina, L: leucina, M: metionina, N: asparagina, P: prolina, Q: glutamina; R: arginina, S: serina, T: treonina, V: valina, W: triptófano, Y: tirosina; X: cualquier residuo). NH₂ hace referencia al grupo amino libre presente en el extremo amino de un polipéptido. COOH hace referencia al grupo carboxilo libre presente en el extremo carboxilo de un polipéptido. Se utiliza la nomenclatura polipeptídica convencional, véase J. Biol. Chem., 243, 3552-59 (1969).

El término “inmunológicamente activo”, define la capacidad de la proteína cromogénica o fluorescente recombinante o sintética, natural, o cualquiera de sus oligopéptidos, para inducir una respuesta inmunitaria específica en animales o células apropiados y para unirse a anticuerpos específicos. Según se utiliza en la presente memoria, “secuencia de aminoácidos antigénica”, significa una secuencia de aminoácidos que, ya sea sola o asociada con una molécula transportadora, puede provocar una respuesta de anticuerpos en un mamífero. El término “unión específica”, en el contexto de la unión de un anticuerpo a un antígeno, es un término bien entendido en la técnica y se refiere a la unión de un anticuerpo al antígeno contra el que fue originado el anticuerpo, pero no a otros, antígenos no relacionados.

Según se utiliza aquí el término “aislado” está destinado a describir un polinucleótido, un polipéptido, un anticuerpo o una célula anfitriona que está en un entorno diferente de aquel en el que el polinucleótido, el polipéptido y el anticuerpo, o la célula anfitriona se produce naturalmente.

La bioluminiscencia es la emisión de luz por organismos vivos que es visible en la oscuridad. (Véase, por ejemplo, Harvey, Bioluminescence, New York: Academic Press (1952); Hastings, “Bioluminescence” in: Cell Physiology (ed. by Seralakis), New York Academic Press págs. 651-81 (1995); Wilson and Hastings, “Bioluminescence”, Annu. Rev. Cell. Dev. Biol. 14, págs.197-230 (1998)). La Bioluminiscencia no incluye la denominada emisión de luz ultradébil, que puede ser detectada en prácticamente todas las estructuras vivientes utilizando un equipo luminométrico sensible (Murphy and Sies, “Visible-range low-level chemiluminescence in biological systems”, Meth. Enzymol. 186, págs. 595-610 (1990); Radotic, *et al.*, “Spontaneous ultra-weak bioluminescence in plants: origin, mechanisms and properties”, Gen. Physiol. Biophys. 17, págs. 289-308 (1998)), ni tampoco emanan bioluminiscencia de la emisión de luz débil que muy probablemente no juega un papel ecológico, tal como la incandescencia de un cono de crecimiento de bambú (Totsune, *et al.*, “Chemiluminescence from bamboo shoot cut”, Biochem. Biophys. Res. Comm. 194, págs. 1025-1029 (1993)), o la emisión de luz durante la fertilización de huevos de animales (Klebanoff, *et al.*, “Metabolic similarities between fertilization and phagocytosis”, J. Exp. Med. 149, págs. 938-53 (1979); Schomer and Epel, “Redox changes during fertilization and maturation of marine invertebrate eggs”, Dev. Biol. 2003, págs. 1-11 (1998)).

Según se utiliza en la presente memoria, el término “proteínas de tipo GFP” está destinado a describir las proteínas similares a la proteína fluorescente verde (GFP) de *Aequorea victoria*.

Las composiciones de ácido nucleico codifican una proteína de tipo GFP incolora, acGFP, de *Aequorea coerulescens*, y sus derivados o mutantes fluorescentes y no fluorescentes, así como las proteínas y péptidos codificados por estos productos compuestos de ácido nucleico. Las proteínas de interés son proteínas que tienen color y/o son fluorescentes y/o pueden ser fotoactivadas, donde la característica de color, fluorescente, o de fotoactivación surge de la interacción de dos o más residuos aminoácido de la proteína. También tienen interés las proteínas que son esencialmente similares a, o los derivados o mutantes de, las proteínas específicas citadas antes. También se proporcionan los fragmentos de ácidos nucleicos y los péptidos codificados por los mismos, así como los anticuerpos para las proteínas y los péptidos sujeto. Además, se proporcionan las células y organismos transgénicos. La proteína sujeto y las composiciones de ácidos nucleicos encuentran uso en una variedad de diferentes aplicaciones y métodos, particularmente en aplicaciones de etiquetado de proteínas. Por último, se proporcionan kits para su uso en tales métodos y aplicaciones.

Antes de describir adicionalmente la invención sujeto, se debe entender que la invención no se limita a las realizaciones concretas de la invención descrita más abajo, puesto que se pueden realizar variaciones de las realizaciones concretas y permanecer todavía dentro del alcance de la invención.

En esta memoria, las formas en singular “un”, “una” “el” y “la” incluyen las referencias en plural a menos que el contexto dicte claramente lo contrario. A menos que se defina de otra manera, todos los términos técnicos y científicos que se utilizan en la presente memoria tienen el mismo significado comúnmente entendido por un experto normal en la técnica a la que pertenece esta invención. Aunque todos los métodos, dispositivos y materiales similares o equivalentes a los descritos en la presente memoria pueden ser utilizados en la práctica o el ensayo de la invención, los métodos, dispositivos y materiales preferidos se describen a continuación.

En la descripción de la presente invención, se describirán en primer lugar las composiciones de ácido nucleico, seguido de una discusión de las composiciones de proteínas, las composiciones de anticuerpos y las células y organismos transgénicos. A continuación, se proporciona un examen de los métodos ilustrativos en los que las proteínas de la presente invención encuentran uso.

Composiciones de Ácidos Nucleicos

Como se ha resumido más arriba, la presente invención proporciona composiciones de ácido nucleico que codifica, una proteína incolora, acGFP, de *Aequorea coerulescens*, o mutantes o derivados fluorescentes de acGFP. La frase “proteína fluorescente” significa una proteína que es fluorescente, por ejemplo, puede presentar fluorescencia baja, media o intensa tras la irradiación con luz de longitud de onda de excitación apropiada. Las proteínas de la presente invención son aquellas en las que la característica fluorescente es la que surge de la interacción de dos o más residuos aminoácido de la proteína, y no de un único residuo aminoácido. Como tales, las proteínas fluorescentes de la presente invención no incluyen proteínas que exhiben fluorescencia sólo a partir de residuos que actúan por sí mismos como fluoróforos intrínsecos, es decir, triptófano, tirosina y fenilalanina. Las proteínas fluorescentes de la invención sujeto por lo tanto son proteínas fluorescentes cuya fluorescencia se origina de alguna estructura de la proteína diferente de los restos aminoácido individuales especificados anteriormente, por ejemplo, se origina a partir de una interacción de dos o más restos aminoácido.

Una composición de ácido nucleico de la presente invención es una composición que comprende una secuencia de ADN con un marco de lectura abierto que codifica un polipéptido de la invención sujeto; es decir, un gen de fluoroproteína. Tal composición de ácido nucleico es capaz, en condiciones apropiadas, de ser expresado como una fluoroproteína. También están abarcados en el término composición de ácido nucleico los ácidos nucleicos que son homólogos a, esencialmente similares a, idénticos a, o miméticos de los ácidos nucleicos de la presente invención. Los ácidos nucleicos sujeto están presentes en un entorno distinto de su entorno natural, por ejemplo, son aislados, están presentes en cantidades enriquecidas, o están presentes o se expresan *in vitro* o en una célula u organismo distinto de su entorno de origen natural.

En otra realización de la presente invención, los ácidos nucleicos pueden ser codificados por los mostrados en las figuras 1, 3, 5, 6, 7, 9, 12, 15, 17, 19, 21 o 23, o los ácidos nucleicos son derivados de, o son homólogos de tales ácidos nucleicos.

Los ácidos nucleicos de la invención sujeto pueden ser ADNc o ADN genómico o uno de sus fragmentos. En algunas realizaciones, los ácidos nucleicos de la invención sujeto incluyen uno o más de los marcos de lectura abiertos que codifican las proteínas y polipéptidos fluorescentes específicos, y los intrones, así como las secuencias de nucleótidos no codificante 5' y 3' adyacentes que participan en la regulación de la expresión, incluyendo secuencias de hasta unos 20 kb más allá de la región codificante, pero posiblemente más, en cualquier dirección. Los ácidos nucleicos sujeto pueden ser introducidos en un vector apropiado para el mantenimiento extracromosómico o para su integración en un genoma anfitrión, como se describe con mayor detalle más abajo.

El término “ADN”, según se utiliza en la presente memoria tiene por objeto incluir los ácidos nucleicos que comparten la disposición de los elementos de la secuencia en especies ARNm maduro nativas, donde los elementos de la secuencia son exones y regiones no codificantes 5' y 3'. Normalmente, las especies de ARNm tienen exones contiguos, eliminándose los intrones intermedios, cuando están presentes, por el empalme del ARN nuclear, para crear un marco de lectura abierto continuo que codifica la proteína.

Una secuencia genómica de interés puede comprender el ácido nucleico presente entre el codón de iniciación y el codón de terminación, como se define en las secuencias enumeradas, incluyendo todos los intrones que normalmente están presentes en un cromosoma nativo. La secuencia genómica de interés puede incluir adicionalmente regiones 5' y 3' no traducidas encontradas en el ARNm maduro, así como secuencias reguladoras transcripcionales y traduccionales específicas, tales como promotores, intensificadores, etc., incluyendo aproximadamente 1 kb, pero posiblemente más, de ADN genómico limítrofe en cualquiera de los extremos 5' o 3' de la región transcrita. El ADN genómico puede ser aislado como un fragmento de 100 kb o menos, y esencialmente exento de secuencias cromosómicas limítrofes. El ADN genómico que flanquea la región codificante, ya sea 3' o 5', o secuencias reguladoras internas como a veces se encuentran en los intrones, puede contener las secuencias necesarias para una expresión apropiada específica del tejido y de la fase.

Los ácidos nucleicos sujeto se pueden aislar y obtener con una pureza sustancial, en general de forma diferente a un cromosoma intacto. Usualmente, el ADN se obtendrá esencialmente libre de secuencias de ácidos nucleicos que no incluyen un ácido nucleico de la invención sujeto o uno de sus fragmentos. Pureza sustancial significa que los ácidos nucleicos son al menos puros en aproximadamente 50%, usualmente al menos puros en aproximadamente 90% y son típicamente “recombinantes”, es decir, están flanqueados por uno o más nucleótidos con los que normalmente no están asociados en un cromosoma de origen natural en su organismo anfitrión natural.

Los polinucleótidos de la presente invención, por ejemplo, los polinucleótidos que tienen la secuencia que se muestra en las Figuras 1, 3, 5, 6, 7, 9, 12, 15, 17, 19, 21 o 23, los ADNc correspondientes, los genes y constructos completos pueden ser generados sintéticamente mediante una serie de protocolos diferentes conocidos por los expertos normales en la técnica. Los constructos de polinucleótidos apropiados son purificados utilizando técnicas de ADN recombinante convencionales como describen, por ejemplo, Sambrook *et al.*, en *Molecular Cloning: A Laboratory Manual*, 2ª ed., Cold Spring Harbor Press, Cold Spring Harbor, NY (1989), y bajo las normas descritas, por ejemplo, en el Departamento HHS de los Estados Unidos, *Instituto Nacional de la Salud (NIH) Directrices para la investigación del ADN recombinante*.

También se proporcionan los ácidos nucleicos que codifican las proteínas de fusión de las proteínas o péptidos sujeto de la presente invención, o sus fragmentos, fusionados a un segundo péptido o proteína. La segunda proteína puede ser, por ejemplo, una secuencia de degradación, un péptido señal, o cualquier proteína de interés. Las proteínas de fusión pueden incluir, por ejemplo, un polipéptido acGFP o acGFP mutante y un segundo polipéptido ("el compañero de fusión") fusionado en marco en el extremo N y/o el extremo C del polipéptido acGFP. Los compañeros de fusión incluyen, pero no están limitados a, polipéptidos que se pueden unir a anticuerpos específicos para el compañero de fusión (por ejemplo, etiquetas epitópicas), anticuerpos o sus fragmentos de unión, polipéptidos que proporcionan una función catalítica o inducen una respuesta celular, sus ligandos o receptores o miméticos, y similares. En tales proteínas de fusión, el compañero de fusión no está generalmente asociado naturalmente a la porción acGFP de la proteína de fusión, y no es típicamente una proteína de *Aequorea coerulescens* o derivado/fragmento de la misma, es decir, no se ha encontrado en especies de *Aequorea*.

También se proporcionan vectores y otros constructos de ácidos nucleicos que comprenden los ácidos nucleicos sujeto, donde tales constructos pueden utilizarse para varias aplicaciones, incluyendo la propagación, la producción de proteínas, etc. Se pueden preparar y utilizar vectores virales y no virales, incluyendo plásmidos. La elección del vector dependerá del tipo de célula en la que se desee la propagación y del propósito de la propagación. Ciertos vectores son útiles para ampliar y elaborar grandes cantidades de la secuencia de ADN deseada. Otros vectores son adecuados para su expresión en células en cultivo. Otros vectores más son adecuados para su transferencia y expresión en las células de un animal completo. La elección del vector adecuado está dentro del conocimiento práctico de la técnica, y muchos de tales vectores están disponibles en el mercado. Para preparar los constructos, el polinucleótido parcial o completo se inserta en un vector típicamente por medio de la unión con ADN ligasa en un sitio escindido por una enzima de restricción en el vector. Alternativamente, la secuencia de nucleótidos deseada puede ser insertada mediante recombinación homóloga *in vivo*, típicamente uniendo las regiones de homología con el vector en los flancos de la secuencia de nucleótidos deseada. Las regiones de homología son añadidas mediante la ligación de oligonucleótidos, o mediante reacción en cadena de la polimerasa utilizando cebadores que comprenden tanto la región de homología como una porción de la secuencia de nucleótidos deseada, por ejemplo.

También se proporcionan casetes o sistemas de expresión que se encuentran en uso, entre otras aplicaciones, en la síntesis de proteínas cromogénicas o fluorescentes o de sus proteínas de fusión. Para su expresión, el producto génico codificado por un polinucleótido de la invención se expresa en cualquier sistema de expresión conveniente, incluyendo, por ejemplo, sistemas bacterianos, de levaduras, insectos, anfibios y mamíferos. Tales vectores y células anfitrionas se describen en la Patente de los Estados Unidos Núm. 5.654.173. En el vector de expresión, un polinucleótido sujeto - por ejemplo, como se muestra en las figuras 1, 3, 5, 6, 7, 9, 12, 15, 17, 19, 21 y 23 - se conecta a una secuencia reguladora según se apropiado para obtener las propiedades de expresión deseadas. Estas secuencias reguladoras pueden incluir promotores (anclados o bien al extremo 5' de la cadena efectora o al extremo 3' de la cadena antisentido), intensificadores, terminadores, operadores, represores e inductores. Los promotores pueden ser regulados o constitutivos. En algunas situaciones puede ser deseable utilizar promotores condicionalmente activos, tales como promotores específicos del tejido o específicos de la fase de desarrollo. Éstos están conectados a las secuencias de nucleótidos deseadas utilizando las técnicas descritas anteriormente para la unión a los vectores. Se puede utilizar cualquier mecanismo conocido en la técnica. En otras palabras, el vector de expresión proporcionará una región de iniciación de la transcripción y la traducción, que puede ser inducible o constitutiva, donde la región codificante está unida operablemente bajo el control transcripcional de la región de la iniciación de la transcripción, y una región de terminación de la transcripción y la traducción. Estas regiones de control pueden ser nativas para la especie sujeto de la que se obtiene el ácido nucleico sujeto, o se pueden obtener de fuentes exógenas.

Los vectores de expresión generalmente tienen sitios de restricción convenientes localizados cerca de la secuencia promotora para proporcionar la inserción de secuencias de ácido nucleico que codifican las proteínas heterólogas. Puede estar presente un marcador seleccionable operativo en el anfitrión de expresión. Se pueden utilizar vectores de expresión para, entre otras cosas, la producción de proteínas de fusión, como se ha descrito anteriormente.

Se pueden preparar casetes de expresión que comprenden una región de iniciación de la transcripción, el gen o su fragmento, y una región de terminación de la transcripción. Tiene particular interés el uso de secuencias que permiten la expresión de epítomos o dominios funcionales, usualmente de al menos aproximadamente 8 aminoácidos de longitud, más usualmente de al menos aproximadamente 15 aminoácidos de longitud, a aproximadamente 25 aminoácidos, y hasta el marco de lectura abierto completo del gen. Después de la introducción del ADN, se pueden seleccionar las células que contienen el constructo por medio de un marcador seleccionable, expandir las células, después utilizarlas para la expresión.

Los sistemas de expresión antes descritos pueden ser empleados con procariotas o eucariotas, de acuerdo con los métodos convencionales, dependiendo de la finalidad de la expresión. Para la producción a gran escala de la proteína, se pueden utilizar un organismo unicelular, tal como *E. coli*, *B. subtilis*, *S. cerevisiae*, células de insecto combinadas con vectores de baculovirus, o células de un organismo superior como los vertebrados, por ejemplo, células COS 7, HEK 293, CHO, oocitos de *Xenopus*, etc., como células anfitrionas de expresión. En algunas situaciones, es deseable expresar el gen en células eucarióticas, donde la proteína expresada se beneficiará de modificaciones de plegamiento y post-transicionales nativas. También se pueden sintetizar péptidos pequeños en el laboratorio. Se pueden utilizar polipéptidos que son subconjuntos de la secuencia de proteínas completa para identificar e investigar las porciones de la proteína importantes para su función.

Los sistemas de expresión específicos de interés incluyen sistemas de expresión derivados de bacterias, levaduras, células de insecto y células de mamíferos. Más abajo se proporcionan las referencias extraídas de los sistemas representativos de cada una de estas categorías.

5 Los sistemas de expresión en bacterias incluyen los descritos por Chang *et al.*, Nature 275:615 (1978); Goeddel *et al.*, Nature 281:544 (1979); Goeddel *et al.*, Nucleic Acids Res. 8:4057 (1980); EP 0 036776; Patente de los Estados Unidos Núm. 4.551.433; DeBoer *et al.*, Proc. Natl. Acad. Sci. (USA) 80:21-25 (1983), Y Siebenlist *et al.*, Cell 20:269 (1980).

10 Los sistemas de expresión en levaduras incluyen los descritos por Hinnen *et al.*, Proc. Natl. Acad. Sci. (USA) 75:1929 (1978); Ito *et al.*, J. Bacteriol. 153:163 (1983); Kurtz *et al.*, Mol. Cell Biol. 6:142 (1986); Kunze *et al.*, J. Basic Microbiol. 25:141 (1985); Gleeson *et al.*, J. Gen. Microbiol. 132:3459 (1986); Roggenkamp *et al.*, Mol. Gen. Genet. 202:302; (1986); Das *et al.*, J. Bacteriol. 158:1165 (1984); De Louvencourt *et al.*, J. Bacteriol. 154:737 (1983); Van den Berg *et al.*, Bio/Technology 8:135 (1990); Kunze *et al.*, J. Basic Microbiol. 25:141 (1985); Cregg *et al.*, Mol. Cell. Biol. 5:3376 (1985); Patentes de los Estados Unidos Núms. 4.837.148 y 4.929.555; Beach and Nurse, Nature 300:706 (1981); Davidow *et al.*, Curr. Genet. 10:380 (1985); Gaillardin *et al.*, Curr. Genet. 10:49 (1985); Ballance *et al.*, Biochem. Biophys. Res. Commun. 112:284-289 (1983); Tilbum *et al.*, Gene 26:205-221 (1983); Yelton *et al.*, Proc. Natl. Acad. Sci. (81:1470-1474 (1984); Kelly and Hynes, EMBO J. 4:475479 (1985); documento EP 0 244, 234: y documento WO 91/00357.

20 La expresión de genes heterólogos en los insectos se realiza como se describe en la Patente de los Estados Unidos Núm. 4.745.051; Friesen *et al.*, "The Regulation of Baculovirus Gene Expression", en: The Molecular Biology of Baculoviruses (W. Doerfler, ed.) (1986); documento EP 0 127, 839; documento EP 0 155, 476; and Vlask *et al.*, J. Gen. Virol. 69:765-776 (1988); Miller *et al.*, Ann. Rev. Microbiol. 42:177 (1988); Carbonell *et al.*, Gene 73:409 (1988); 25 Maeda *et al.*, Nature 315:592-594 (1985); Labacq-Verheyden *et al.*, Mol. Cell. Biol. 8:3129 (1988); Smith *et al.*, Proc. Natl. Acad. Sci. 82:8844 (1985); Miyajima *et al.*, Gene 58:273 (1987); and Martin *et al.*, DNA 7:99 (1988). Numerosas cepas y variantes baculovirales y las correspondientes células anfitrionas de insectos permisivas de los anfitriones son descritas por Luckow *et al.*, Bio/Technology 6:47-55 (1988); Miller *et al.*, Genetic Engineering 8:277-279 (1986); and Maeda *et al.*, Nature 315:592-594 (1985).

30 La expresión en mamíferos se realiza como describen Dijkema *et al.*, EMBO J. 4:761 (1985), Gorman *et al.*, Proc. Natl. Acad. Sci. (EE.UU.) 79:6777 (1982); Boshart *et al.*, Cell 41:521 (1985), y la Patente de los Estados Unidos Núm. 4.399.216. Otras características de la expresión en mamíferos son facilitadas como describen Hamm y Wallace, Meth. Enz. 58:44 (1979); Barnes y Sata, Anal. Biochem. 102:255 (1980); Patentes de los Estados Unidos Núms. 35 4.767.704; 4.657.866; 4.927.762; 4.560.655; documento WO 90/103430; documento WO 87/00195, y documento U.S. RE 30.985.

Cuando se utilizan cualquiera de las células anfitrionas antes mencionados, u otras células u organismos anfitriones apropiados para reproducir y/o expresar los polinucleótidos o ácidos nucleicos de la invención, el ácido nucleico replicado, el ARN, las proteínas o el polipéptido expresados resultantes se encuentran dentro del alcance de la invención como producto de la célula u organismo anfitriones. El producto puede ser recuperado mediante un dispositivo apropiado conocido en la técnica.

Una vez que se identifica el gen correspondiente a un polinucleótido seleccionado, su expresión puede ser regulada en la célula de la que es nativo el gen. Por ejemplo, un gen endógeno de una célula puede ser regulado por una secuencia reguladora exógena insertada en el genoma de la célula en la ubicación suficiente para al menos aumentar la expresión del gen en la célula. La secuencia reguladora se puede diseñar para integrarla en el genoma a través de la recombinación homóloga, como se describe en las Patentes de los Estados Unidos Núms. 5.641.670 y 5.733.761, cuyas descripciones son incorporadas en la presente memoria como referencia, o se puede diseñar para integrarla en el genoma a través de recombinación no homóloga, como se describe en el documento WO 99/15650, cuya descripción se incorpora también en la presente memoria como referencia. Como tal, también se incluye en la presente invención la producción de proteínas sin manipulación del propio ácido nucleico codificante, integrando en su lugar una secuencia reguladora en el genoma de la célula que ya incluye un gen que codifica la proteína deseada.

55 También tienen interés las secuencias promotoras de las secuencias genómicas de la presente invención, donde la secuencia de la región limítrofe 5' se puede utilizar para los elementos promotores, incluyendo los sitios de unión al intensificador, que, por ejemplo, proporcionan la regulación de la expresión en células/tejidos donde se expresan los genes de las proteínas sujeto.

60 La composición de ácidos nucleicos de la presente invención también se puede utilizar para identificar la expresión de un gen en un espécimen biológico. La manera en que se sondean las células para detectar la presencia de secuencias de nucleótidos concretas, por ejemplo en ADN o ARN genómico, está bien establecida en la técnica. En pocas palabras, el ADN o ARNm se aíslan de una muestra celular. El ARNm puede ser amplificado por RT-PCR, usando la transcriptasa inversa para formar una cadena complementaria de ADN, seguido de la amplificación mediante reacción en cadena de la polimerasa utilizando cebadores específicos para las secuencias sujeto de ADN. Alternativamente, la muestra de ARNm se separa mediante electroforesis en gel, se transfiere a un soporte adecuado, por ejemplo, nitrocelulosa, nailon, etc., y luego se sondea con un fragmento del ADN sujeto como sonda. También se pueden utilizar

otras técnicas, tales como los análisis de ligación de oligonucleótidos, hibridaciones *in situ*, e hibridación con sondas de ADN dispuestas en un chip sólido. La detección de ARNm que hibrida con la secuencia sujeto es indicativa de la expresión génica en la muestra.

5 Composiciones de Péptidos

La invención sujeto proporciona la proteína acGFP fluorescente. Según se utiliza en la presente memoria, el término proteína fluorescente se refiere a cualquier proteína que se ilumina cuando se irradia con luz, por ejemplo, luz blanca o luz de una longitud de onda específica (o una banda estrecha de longitudes de onda, tal como una longitud de onda de excitación). El término polipéptido utilizado en la presente invención se refiere tanto a las proteínas completas, como a porciones o fragmentos de proteínas. También se incluyen en este término las variaciones de una proteína de origen natural, donde tales variaciones son homólogas o esencialmente similares a la proteína de origen natural, y los mutantes de las proteínas de origen natural, como se describe con mayor detalle a continuación. Los polipéptidos sujeto están presentes en otros entornos distintos de su entorno natural.

En muchas realizaciones, las proteínas sujeto tienen un máximo de absorbancia de 300 nm a 700 nm, usualmente de aproximadamente 350 nm a 550 nm y más usualmente de aproximadamente 400 a 500 nm, y con frecuencia de aproximadamente 450 a 490 nm, por ejemplo, de 470 a 490 nm, mientras que los espectros de emisión de las proteínas sujeto oscila típicamente de aproximadamente 400 nm a 700 nm, usualmente de aproximadamente 450 nm a 650 nm y más usualmente de aproximadamente 500 a 600 nm, mientras que en muchas realizaciones los espectros de emisión oscilan de aproximadamente 500 a 550 nm, por ejemplo, 500 a 525 nm, o de 500 a 510 nm. Las proteínas sujeto generalmente tienen un coeficiente de extinción máximo que oscila entre 25.000 y 150.000 y usualmente de aproximadamente 45.000 a 120.000, por ejemplo, de 50.000 a 100.000. Las proteínas sujeto varían típicamente de longitud desde aproximadamente 150 a 300 aminoácidos y usualmente de aproximadamente 200 a 300 residuos aminoácido y, generalmente tienen un peso molecular que oscila de aproximadamente 15 a 35 kDa, usualmente de aproximadamente 17,5 a 32,5 kDa.

En algunas realizaciones, las proteínas sujeto son luminosas, donde “luminosas” quiere decir que las cromoproteínas y sus derivados fluorescentes se pueden detectar mediante métodos comunes (por ejemplo, detección visual, espectrofotometría, espectrofluorometría, microscopia de fluorescencia, mediante instrumentación FACS, etc.). La luminosidad de la fluorescencia de una proteína fluorescente concreta está determinada por su rendimiento cuántico, multiplicado por el coeficiente de extinción máximo. La luminosidad de una cromoproteína puede ser expresada por su coeficiente de extinción máximo.

En algunas realizaciones, las proteínas sujeto se pliegan rápidamente después de su expresión en la célula anfitriona. “Plegadas rápidamente” significa que las proteínas alcanzan la estructura terciaria que da lugar a su cualidad cromogénica o fluorescente en un período corto de tiempo. En estas realizaciones, las proteínas se pliegan en un período de tiempo que generalmente no supera los 3 días, usualmente no supera los 2 días y más generalmente no supera 1 día.

Las proteínas específicas de interés incluyen la proteína fluorescente acGFP de tipo salvaje y sus mutantes, según se proporciona, por ejemplo, en las Figuras 1, 3, 5, 6, 7, 9, 12, 15, 17, 19, 21 o 23, que codifican acGFP, Z1, Z2, G1, G2, G22, G22E, G22-G22E/Y220L, 220-11-5, CFP-rand3 y CFP-3 y G22 humanizada.

También se proporcionan los homólogos de proteínas cuyas secuencias difieren de las secuencias de aminoácidos específicas proporcionadas anteriormente, es decir, las Figuras 1, 3, 5, 6, 7, 9, 12, 15, 17, 19, 21 o 23. “Homóloga” en relación con la invención significa una proteína que tiene al menos 95% de identidad de secuencia de aminoácidos con las secuencias de aminoácidos que se muestran en las Figuras 1, 3, 5, 6, 7, 9, 12, 15, 17, 19, 21 o 23, según se determina utilizando MegAlign, el algoritmo de DNASTAR Clustal descrito por D.G. Higgins y P.M. Sharp, “Fast and Sensitive multiple Sequence Alignments on a Microcomputer”. CABIOS, 5 págs. 151-3 (1989) (utilizando parámetros ktuple 1, penalización del espacio 3, ventana 5 y diagonales salvadas 5).

También se proporcionan las proteínas que son esencialmente idénticas a la proteína de tipo salvaje, donde “esencialmente idéntica” significa que la proteína tiene una identidad de secuencia de aminoácidos con respecto a la secuencia de la proteína de tipo salvaje de al menos 95% o superior, por ejemplo, 95%, 96%, 97%, 98%, 99%, 99,5%.

También se proporcionan las proteínas que son derivados o mutantes de las proteínas de origen natural antes descritas. Los mutantes pueden conservar las propiedades biológicas de las proteínas tipo salvaje (por ejemplo, de origen natural), o pueden tener propiedades biológicas que difieren de las proteínas de tipo salvaje. El término “propiedad biológica” de las proteínas de la presente invención se refiere, pero no está limitado a, las propiedades espectrales, tales como el máximo de absorbancia, el máximo de emisión, el coeficiente de extinción máximo, la luminosidad (por ejemplo, en comparación con la proteína de tipo salvaje u otra proteína de referencia, tal como la proteína fluorescente verde (GFP) de *A. Victoria*), y similares; la estabilidad *in vivo* y/o *in vitro* (por ejemplo, vida media); y otras propiedades semejantes. Las mutaciones incluyen cambios de un solo aminoácido, delecciones de uno o más aminoácidos, truncamientos N-terminales, truncamientos C-terminales, inserciones, y similares.

Las proteínas mutantes pueden ser generadas utilizando técnicas convencionales de biología molecular, por ejemplo, mutagénesis aleatoria, y mutagénesis dirigida como se ha descrito anteriormente. En la presente memoria se describen varios mutantes. Teniendo en cuenta la orientación proporcionada en el Ejemplo, y utilizando técnicas

convencionales, los expertos en la técnica pueden generar fácilmente una gran variedad de mutantes adicionales y comprobar si una propiedad biológica ha sido alterada. Por ejemplo, la intensidad de la fluorescencia se puede medir utilizando un espectrofotómetro a diferentes longitudes de onda de excitación.

5 Esas proteínas de la invención sujeto que son proteínas de origen natural están presentes en un entorno de origen no natural, por ejemplo, son separadas de su entorno de origen natural. En algunas realizaciones, las proteínas sujeto están presentes en una composición que se enriquece con la proteína sujeto en comparación con su entorno de origen natural. Por ejemplo, se proporciona la proteína purificada, donde “purificada” significa que la proteína está presente en una composición que está esencialmente libre de: proteínas no cromogénicas o fluorescentes de interés, donde
10 “esencialmente libre” significa que menos de 90%, usualmente menos de 60% y más usualmente menos de 50% de la composición está formada por proteínas no cromogénicas o fluorescentes o sus mutantes. Las proteínas de la presente invención también pueden estar presentes como un producto aislado, por lo que se entiende que la proteína está esencialmente libre de otras proteínas y otras moléculas biológicas de origen natural, tales como oligosacáridos, polinucleótidos y sus fragmentos, y similares, donde el término “esencialmente libre” en este caso significa que menos
15 de 70%, usualmente menos del 60% y más usualmente menos de 50% de la composición que contiene la proteína aislada es cualquier otra molécula biológica de origen natural. En algunas realizaciones, las proteínas están presentes en forma esencialmente pura, donde “forma esencialmente pura” significa al menos 95%, usualmente al menos 97% y más generalmente al menos 99% pura.

20 Además de las proteínas que de origen natural, también se proporcionan polipéptidos que varían de las proteínas de origen natural, por ejemplo, las proteínas mutantes descritas anteriormente. Generalmente tales polipéptidos incluyen una secuencia de aminoácidos codificada por un marco de lectura abierto (ORF) del gen que codifica la proteína sujeto de tipo salvaje, incluyendo la proteína completa y sus fragmentos, particularmente fragmentos biológicamente activos y/o fragmentos que corresponden a dominios funcionales, y similares; incluyendo fusiones de los polipéptidos sujeto
25 a otras proteínas o péptidos.

Las proteínas y polipéptidos sujeto pueden obtenerse de fuentes de origen natural o producirse sintéticamente. Por ejemplo, las proteínas de tipo salvaje pueden obtenerse de fuentes biológicas que expresan las proteínas, por ejemplo, *Aequorea coerulescens*; las proteínas sujeto también pueden obtenerse a partir de medios sintéticos, por
30 ejemplo, mediante la expresión de una secuencia codificante de un gen o de un ácido nucleico recombinante que codifica la proteína de interés en un organismo adecuado, como se ha descrito anteriormente. Se pueden emplear otros procedimientos de purificación de proteínas convenientes, donde las metodologías de purificación de proteínas adecuadas se describen en Guide to Protein Purification, (Deuthser ed.) (Academic Press, 1990). Por ejemplo, se puede preparar un producto lisado a partir de la fuente original y purificar mediante HPLC, cromatografía de exclusión,
35 electroforesis en gel, cromatografía de afinidad, y similares.

Composiciones de Anticuerpos

También se proporcionan anticuerpos que se unen específicamente a las proteínas fluorescentes de la presente invención. Los anticuerpos adecuados se obtienen mediante la inmunización de un animal anfitrión con péptidos que comprenden la totalidad o una porción de la proteína. Los animales anfitriones adecuados incluyen ratones, ratas, ovejas, cabras, hámsteres, conejos, y otros. El inmunógeno puede comprender la proteína completa, o sus fragmentos y derivados. Los inmunógenos preferidos comprenden la totalidad o una parte de la proteína, donde la proteína incluye las modificaciones post-traduccionales encontradas en la proteína objetivo nativa. Los inmunógenos se producen en
45 una variedad de formas conocidas en la técnica, por ejemplo, expresión de genes clonados utilizando los métodos recombinantes convencionales, o aislamiento directamente a partir de *Aequorea coerulescens*.

Para la preparación de anticuerpos policlonales, el primer paso implica la inmunización de los animales anfitriones con el inmunógeno peptídico, donde inmunógeno proteico peptídico estará preferiblemente en forma esencialmente
50 pura, comprendiendo menos de aproximadamente 1% de contaminantes. El inmunógeno puede comprender una proteína completa, o sus fragmentos o derivados. Para aumentar la respuesta inmunitaria del animal anfitrión, la proteína diana se puede combinar con un coadyuvante, donde los coadyuvantes adecuados incluyen alumbre, dextrano, sulfato, grandes aniones poliméricos, emulsiones de aceite y agua, coadyuvante de Freund, coadyuvante completo de Freund, y similares. El inmunógeno peptídico también se puede conjugar con proteínas portadoras sintéticas o antígenos sin-
55 téticos.

El inmunógeno peptídico es administrado al anfitrión, usualmente por vía intradérmica, con una dosis inicial seguida por una o más, usualmente al menos dos dosis de refuerzo adicionales. Después de la inmunización, se recogerá la sangre del anfitrión, donde el suero sanguíneo se separa de las células sanguíneas. La inmunoglobulina presente en
60 el antisuero resultante se puede purificar utilizando métodos conocidos, tales como el fraccionamiento con la sal de amonio, cromatografía de DEAE, y similares.

Alternativamente, los anticuerpos monoclonales se pueden producir mediante técnicas convencionales. En general, el bazo o los ganglios linfáticos de un animal anfitrión inmunizado proporcionan una fuente de células plasmáticas. Las
65 células plasmáticas son inmortalizadas mediante la fusión con células de mieloma para producir células de hibridoma. El sobrenadante de cultivo de los hibridomas individuales se escruta utilizando técnicas convencionales para identificar los que producen anticuerpos con la especificidad deseada. Los animales aptos para la producción de anticuerpos monoclonales para la proteína humana incluyen ratones, ratas, hámsters, y similares. Para crear anticuerpos contra

la proteína de ratón, el animal generalmente, será un hamster, cobaya o conejo. El anticuerpo puede ser purificado a partir de los sobrenadantes celulares de los hibridomas o del líquido ascítico mediante técnicas convencionales, como cromatografía de afinidad utilizando proteína unida a un soporte insoluble como proteína A-sefarosa.

- 5 El anticuerpo puede ser producido en forma de una sola cadena, en lugar de la estructura normal de multímeros. Los anticuerpos de cadena sencilla son descritos por Jost *et al.*, J.B.C. 269:26267-73 (1994), y otros. Las secuencias de ADN que codifican la región variable de la cadena pesada y de la región variable de la cadena ligera se pueden ligar a un espaciador que codifica al menos aproximadamente 4 aminoácidos de los aminoácidos neutros pequeños, incluyendo glicina y/o serina. La proteína codificada por esta fusión permite el ensamblaje de una región variable
10 funcional que conserva la especificidad y afinidad del anticuerpo original.

- También tienen interés en ciertas realizaciones los anticuerpos humanizados. Los métodos de humanización de anticuerpos son conocidos en la técnica. El anticuerpo humanizado puede ser el producto de un animal que tiene genes de la región constante de la inmunoglobulina humana transgénica (véanse por ejemplo, las Solicitudes de Patentes
15 Internacionales WO 90/10077 y WO 90/04036). Alternativamente, el anticuerpo de interés puede ser diseñado mediante técnicas de ADN recombinante para sustituir los dominios bisagra CH1, CH2, CH3, y/o el dominio marco con la secuencia humana correspondiente (véase el documento WO 92/02190).

- El uso de ADNc de inmunoglobulina para la construcción de los genes de la inmunoglobulina quimérica es conocido
20 en la técnica (Liu *et al.*, Actas de la Academia Nacional de Ciencias 84:3439 (1987) Y J. Immunol. 139:3521 (1987)). Esencialmente, el ARNm es aislado de un hibridoma u otra células que produzca los anticuerpos y se utiliza para producir ADNc. El ADNc de interés puede ser amplificado mediante la reacción en cadena de la polimerasa con iniciadores específicos (Patente de los Estados Unidos Núms. 4.683.195 y 4.683.202). Alternativamente, se elabora y se escruta una genoteca para aislar la secuencia de interés. La secuencia de ADN que codifica la región variable
25 de los anticuerpos se fusiona después con las secuencias de la región constante humana. Las secuencias de los genes de las regiones constantes humanas se pueden encontrar en Kabat *et al.* "Sequences of Proteins of Immunological Interest", N.I.H. publication no. 91-3242 (1991). Los genes de la región C humana están disponibles a partir de clones conocidos. La elección del isotipo se guiará por las funciones efectoras deseadas, tales como la fijación del complemento, o la actividad en la citotoxicidad celular dependiente del anticuerpo. Los isotipos preferidos son IgG1,
30 IgG3 e IgG4. Se puede utilizar cualquiera de las regiones constantes de la cadena ligera humana, kappa o lambda. El anticuerpo humanizado quimérico se expresa después mediante métodos convencionales.

- Los fragmentos de anticuerpo, como Fv, F(ab')₂ y Fab se pueden preparar mediante la escisión de la proteína intacta, por ejemplo, mediante una proteasa o escisión química. Alternativamente, se puede diseñar un gen truncado.
35 Por ejemplo, un gen quimérico que codifica una porción del fragmento F(ab')₂ indicaría las secuencias de ADN que codifican el dominio CH1 y región bisagra de la cadena H, seguido por un codón de terminación de la traducción para producir la molécula truncada.

- Se pueden utilizar secuencias consenso de las regiones H y L J para diseñar oligonucleótidos para su uso como
40 cebadores para introducir sitios de restricción útiles en la región de J para la posterior unión de los segmentos de la región V a los segmentos de la región C humana. La región C del ADNc puede ser modificada mediante mutagénesis dirigida al sitio para establecer un sitio de restricción en la posición análoga en la secuencia humana.

- Los vectores de expresión incluyen plásmidos, retrovirus, YAC, episomas derivados de EBV, y similares. Un vector
45 conveniente es el que codifica una secuencia de inmunoglobulina CH o CL humana funcionalmente completa, con sitios de restricción apropiados diseñados de modo que se pueda insertar y expresar fácilmente cualquier secuencia VL o VH. En tales vectores, el empalme se produce usualmente entre el sitio donador de empalme en la región J insertada y el sitio aceptor de empalme que precede a la región C humana, y también en las regiones de empalme que se producen dentro de los exones CH humanos. La poliadenilación y la terminación de la transcripción se producen
50 en sitios cromosómicos nativos aguas abajo de las regiones codificantes. El anticuerpo quimérico resultante se puede unir a cualquier promotor fuerte, incluyendo las LTR antirretrovirales, tales como el promotor temprano de SV-40, (Okayama *et al.*, Mol. Cell. Bio. 3:280 (1983)); la LTR del virus del sarcoma de Rous (Gorman *et al.*, Proceedings of the National Academy of Sciences. 79:6777 (1982)); o la LTR del virus de la leucemia murina de Moloney (Grosschedl
et al., Cell 41:885 (1985)); o promotores de Ig nativos.

55 Transgénicos

- Los ácidos nucleicos de la presente invención se pueden utilizar para generar las plantas o animales no humanos transgénicos, o modificaciones génicas específicas del sitio en las líneas celulares. Las células transgénicas de la
60 invención sujeto incluyen uno o más ácidos nucleicos de acuerdo con la invención sujeto presente en forma de transgén, estando incluidas en esta definición las células parentales transformadas para incluir el transgén y su progenie. En muchas realizaciones, las células transgénicas son células que no albergan o contienen normalmente un ácido nucleico de acuerdo con la presente invención. En las realizaciones en las que las células transgénicas contienen naturalmente los ácidos nucleicos sujeto, el ácido nucleico estará presente en la célula en una posición distinta de su localización
65 natural, estando integrado por ejemplo en el material genómico de la célula en su localización no natural. Los animales transgénicos no humanos pueden ser realizados a través de recombinación homóloga, cuando se modifica el locus endógeno. Alternativamente, se integra de forma aleatoria en el genoma un constructo de ácido nucleico. Los vectores para la integración estable incluyen plásmidos, retrovirus y otros virus animales, YAC, y similares.

Los organismos transgénicos no humanos de interés incluyen células y organismos multicelulares, tanto plantas como animales no humanos, en los que la proteína o sus variantes se expresan en células o tejidos en los que no se expresan normalmente y/o a niveles no normalmente presentes en tales células o tejidos.

5 Los constructos de ADN para su recombinación homóloga comprenderán al menos una porción de un ácido nucleico de la presente invención, donde el gen tiene la modificación o modificaciones genéticas deseadas, e incluye las regiones de homología con el locus diana. Los constructos de ADN para la integración al azar no necesitan incluir las regiones de homología para mediar la recombinación. Convenientemente, se pueden incluir los marcadores para la selección positiva y negativa. Los métodos para la generación de células que tienen modificaciones génicas dirigidas a través de recombinación homóloga son conocidos en la técnica. Para las diferentes técnicas para transfectar células de mamíferos, véase Keown *et al.*, Meth. Enzymol. 185:527-37 (1990).

15 Para las células madre embrionarias (ES), se puede emplear una línea celular ES no humana, o se pueden obtener células embrionarias de nueva aportación de un anfitrión, tal como un ratón, rata, cobaya, etc. Tales células se hacen crecer en una capa de alimentación de fibroblastos apropiada o se hacen crecer en presencia del factor inhibidor de leucemia (LIF). Cuando las ES o células embrionarias se han transformado, pueden ser utilizadas para producir animales no humanos transgénicos. Después de la transformación, las células se cultivan en placa en una capa de alimentación en un medio apropiado. Las células que contienen el constructo se pueden detectar empleando un medio selectivo. Después de dar el tiempo suficiente para que crezcan las colonias, las colonias se recogen y se analizan en busca de la aparición de recombinación homóloga o integración del constructo. Las colonias que son positivas podrán utilizarse después para la manipulación de embriones y la inyección de blastocistos. Los blastocistos son obtenidos de hembras superovuladas de 4 a 6 semanas de edad. Las células ES se tratan con tripsina, y las células modificadas se inyectan en el blastocelo del blastocisto. Después de la inyección, los blastocitos se devuelven a cada cuerno uterino de las hembras pseudogestantes. Después se deja que las hembras lleguen a término y se escruta la descendencia resultante en busca del constructo. La progenie quimérica se puede detectar fácilmente si el fenotipo de las células transformadas difiere de algún modo de las células de origen natural (por ejemplo exhibiendo fluorescencia).

30 Los animales quiméricos son examinados para detectar la presencia del gen modificado y los machos y hembras que tienen la modificación se emparejan para producir descendencia homocigótica. Si las alteraciones de los genes causan letalidad en algún momento en el desarrollo, como es posible en particular con las proteínas de fusión de la presente invención, los tejidos u órganos pueden mantenerse como injertos o trasplantes alogénicos o congénicos, o en cultivo *in vitro*. Los animales transgénicos pueden ser mamíferos no humanos cualesquiera, tales como animales de laboratorio, animales domésticos, etc., y se utilizan en estudios funcionales, detección de drogas y similares. Los ejemplos representativos de la utilización de animales transgénicos incluyen los descritos más abajo.

35 También se pueden producir plantas transgénicas. Los métodos de preparación de células vegetales y plantas transgénicas se describen en las Patentes de los Estados Unidos Núms. 5.767.367; 5.750.870; 5.739.409; 5.689.049; 5.689.045; 5.674.731; 5.656.466; 5.633.155; 5.629.470; 5.595.896; 5.576.198; 5.538.879; 5.484.956; cuyas descripciones se incorporan en la presente memoria como referencia. Los métodos de producción de plantas transgénicas también se examinan en Plant Biochemistry and Molecular Biology (eds. Lea and Leegood, John Wiley & Sons) págs. 275-295 (1993). En resumen, se recogen una célula o tejido vegetal adecuados, dependiendo de la naturaleza de las especies de plantas. Como tales, en ciertos casos, se aislarán protoplastos, donde tales protoplastos se pueden aislar de una variedad de tejidos vegetales diferentes, por ejemplo, hoja, hipocotilo, raíz, etc. Para el aislamiento de los protoplastos, las células recolectadas se incuban en presencia de celulasas con el fin de eliminar la pared celular, donde las condiciones exactas de incubación varían en función del tipo de planta y/o tejido del que derive la célula. Los protoplastos resultantes se separan después de los restos celulares resultantes mediante tamizado y centrifugación.

50 Alternativamente, los explantes embriogénicos que comprenden las células somáticas se pueden utilizar para la preparación del anfitrión transgénico. Después de la recolección de células o tejidos, el ADN exógeno de interés se introduce en las células vegetales, donde una variedad de diferentes técnicas está disponible para su introducción. Con los protoplastos aislados, se presenta la oportunidad para la introducción a través de protocolos de transferencia génica mediada por ADN, incluyendo la incubación de los protoplastos con ADN desnudo, por ejemplo los plásmidos que comprenden la secuencia codificante exógena de interés en presencia de cationes polivalentes (por ejemplo, PEG o PLO); o la electroporación de los protoplastos en presencia de ADN desnudo que comprende la secuencia exógena de interés. Los protoplastos que han absorbido con éxito el ADN exógeno se seleccionan después, se desarrollan hasta un callo, y finalmente hasta una planta transgénica a través del contacto con las cantidades y proporciones de factores estimuladores apropiadas, tales como auxinas y citoquininas.

60 Con los explantes de embriones, un método conveniente de introducir el ADN exógeno en las células somáticas diana es a través del uso de protocolos de aceleración de partículas o "pistola génica". Los explantes resultantes se dejan crecer después hasta las plantas quiméricas, se cruzan, y después se obtiene la progenie transgénica.

65 En lugar de los enfoques con ADN desnudo descritos anteriormente, otro método de producción de plantas transgénicas es a través de la transformación mediada por *Agrobacterium*. Con la transformación mediada por *Agrobacterium*, se preparan vectores de co-integración o binarios que comprenden el ADN exógeno y después se introducen en una cepa de *Agrobacterium* apropiada, por ejemplo, *A. tumefaciens*. La bacteria resultante se incuba después con protoplastos o explantes de tejido preparados, tales como un disco de hoja, y se produce un callo. El callo se cultiva después

en condiciones selectivas, se selecciona y se somete a los medios de crecimiento para inducir el crecimiento de la raíz y los brotes para producir finalmente una planta transgénica.

Métodos de uso

Las proteínas y péptidos fluorescentes de la presente invención encuentran uso en una variedad de aplicaciones diferentes. Los usos representativos de cada uno de estos tipos de proteínas se describen a continuación, donde los usos descritos en la presente memoria son meramente ejemplares y no tienen ninguna intención de limitar el uso de las proteínas de la presente invención a los descritos.

La primera aplicación de interés es el uso de las proteínas sujeto en métodos de transferencia de energía resonante de fluorescencia (FRET). En estos métodos, las proteínas sujeto sirven como donadores y/o aceptores combinadas con una segunda proteína o colorante fluorescente, por ejemplo, una proteína fluorescente como describen Matz *et al.*, Nature Biotechnology 17:969-973 (1999); una proteína fluorescente verde de *Aequorea victoria* o sus mutantes fluorescentes, por ejemplo, como se describe en las Patentes de los Estados Unidos Núms. 6.066.476; 6.020.192; 5.985.577; 5.976.796; 5.968.750; 5.968.738; 5.958.713; 5.919.445; 5.874.304, cuyas descripciones se incorporan a la presente memoria como referencia; otros colorantes fluorescentes tales como la cumarina y sus derivados, 7-amino-4-metilcumarina y aminocumarina; colorantes bodipy; cascada azul; o fluoresceína y sus derivados, tales como isotiocianato de fluoresceína y verde Oregón; colorantes de rodamina tales como rojo de Texas, tetrametilrodamina, eosinas y eritrosinas; colorantes de cianina tales como Cy3 y Cy5; quelatos macrocíclicos de iones lantánidos, tales como colorante quantum, y colorantes quimioluminiscentes tales como luciferasas, incluyendo las descritas en las Patentes de los Estados Unidos Núms. 5.843.746; 5.700.673; 5.674.713; 5.618.722; 5.418.155; 5.330.906; 5.229.285; 5.221.623; 5.182.202; cuyas descripciones se incorporan a la presente memoria como referencia.

Los ejemplos específicos en los que se pueden utilizar los análisis FRET que emplean las proteínas fluorescentes sujeto incluyen, pero están limitados a, la detección de interacciones proteína-proteína, tal como en un sistema de dos híbridos de mamíferos, la dimerización del factor de transcripción, la multimerización de proteínas de membrana, la formación de complejos multiprotéicos; tal como un biosensor para varios eventos diferentes, donde un péptido o proteína une covalentemente una combinación fluorescente FRET que incluye las proteínas fluorescentes sujeto y el péptido o proteína conectores son, por ejemplo, un sustrato específico de la proteasa para la escisión mediada por caspasa, un péptido que sufre un cambio conformacional al recibir una señal que aumenta o disminuye FRET, tal como un dominio regulador PKA (AMPC sensor), un sitio de fosforilación (por ejemplo, donde hay un sitio de fosforilación en el péptido o el péptido tiene especificidad de unión al dominio fosforilado/desfosforilado de otra proteína), o el péptido tiene dominio de unión a Ca^{2+} . Además, las aplicaciones de transferencia de energía resonante de fluorescencia o FRET en las que las proteínas de la presente invención encuentran uso incluyen, pero no están limitadas a, las descritas en: las Patentes de los Estados Unidos Núms. 6.008.373; 5.998.146; 5.981.200; 5.945.526; 5.945.283; 5.911.952; 5.869.255; 5.866.336; 5.863.727; 5.728.528; 5.707.804; 5.688.648; 5.439.797; cuyas descripciones se incorporan a la presente memoria como referencia.

Las proteínas fluorescentes sujeto también encuentran uso como biosensores en células procarióticas y eucarióticas, por ejemplo como indicadores de ión Ca^{2+} ; indicadores de pH, indicadores de fosforilación, o como indicadores de otros iones, tales como magnesio, sodio, potasio, cloruro y haluros. Por ejemplo, para la detección de iones Ca^{2+} , se sabe que las proteínas que contienen un motivo en mano EF se traslocan desde el citosol a las membranas después de la unión a Ca^{2+} . Estas proteínas contienen un grupo miristoilo que está oculto dentro de la molécula por interacciones hidrofóbicas con otras regiones de la proteína. La unión de Ca^{2+} induce un cambio conformacional exponiendo el grupo miristoilo que a continuación se encuentra disponible para la inserción en la bicapa lipídica (denominado "interruptor de Ca^{2+} -miristoilo"). La fusión de tal proteína que contiene la mano EF con las proteínas fluorescentes la convertirá en un indicador del Ca^{2+} intracelular por el control de la traslocación desde el citosol a la membrana plasmática mediante microscopía confocal. Las proteínas en mano EF adecuadas para su uso en este sistema incluyen, pero no están limitadas a: recoverina, calcineurina B, troponina C, visinina, neurocalcina, calmodulina, parvalbúmina, y similares.

Para indicar el pH, se puede emplear un sistema basado en hisactofilinas. Se sabe que las hisactofilinas son proteínas ricas en histidina miristoilada que existen en *Dictyostelium*. Su unión a la actina y a los lípidos ácidos es fuertemente dependiente del pH dentro del rango de las variaciones de pH del citoplasma. En las células vivas, la unión a la membrana parece reemplazar la interacción de las isactofilinas con los filamentos de actina. A $\text{pH} \leq 6,5$ se localizan en la membrana plasmática y el núcleo. En cambio, a un pH de 7,5 se distribuyen uniformemente por todo el espacio del citoplasma. Este cambio de distribución es reversible y se atribuye a agrupamientos de histidina expuestos en bucles en la superficie de la molécula. La reversión de la distribución intracelular en el intervalo de variaciones de pH del citoplasma está de acuerdo con un pK de 6,5 de los residuos de histidina. La distribución celular es independiente de la miristoilación de la proteína. Mediante la fusión de las proteínas fluorescentes a la hisactofilina, la distribución intracelular de la proteína de fusión se puede seguir por barrido láser, microscopía confocal y microscopía de fluorescencia convencional.

Para tales estudios, el análisis cuantitativo de la fluorescencia se puede hacer mediante la realización de exploraciones en línea a través de las células (microscopía confocal de barrido láser) u otros análisis de datos electrónicos (por ejemplo, utilizando el soporte lógico Metamorph (Universal Imaging Corp)) y un promedio de los datos recogidos en una población de células. La redistribución esencial dependiente de pH de la hisactofilina/proteína fluorescente del citosol a la membrana plasmática se produce en 1-2 minutos y alcanza un nivel de estado estacionario después de 5-10

minutos. La reacción inversa tiene lugar en una escala de tiempo similar. Como tal, se puede utilizar una proteína de fusión hisactofilina/proteína fluorescente que actúa de forma análoga para verificar los cambios del pH citosólico en tiempo real en células de mamíferos vivos. Estos métodos tienen uso en aplicaciones de alto rendimiento, por ejemplo, en la medición de los cambios de pH como consecuencia de la activación del receptor del factor de crecimiento (por ejemplo, factor de crecimiento epitelial o derivado de plaquetas), estimulación quimiotáctica/locomoción celular, en la detección de cambios de pH intracelular como segundo mensajero, en el control de pH intracelular en experimentos de manipulación del pH, y similares.

Para la detección de la actividad de la PKC, el sistema informador aprovecha el hecho de que una molécula llamada MARCKS (sustrato de quinasa C rico en alanina miristoilada) es un sustrato de la PKC. MARCKS está anclado a la membrana plasmática vía miristoilación y un tramo de aminoácidos cargados positivamente (dominio ED) que se une a la membrana plasmática cargada negativamente a través de interacciones electrostáticas. Tras la activación de la PKC, el dominio ED acaba fosforilado por la PKC, cargándose así negativamente, y como consecuencia de la repulsión electrostática MARCKS se trasloca de la membrana plasmática al citoplasma (denominado “interruptor electrostático de miristoilo”). La fusión del extremo N del motivo de la miristoilación de MARCKS al dominio ED de MARCKS de las proteínas fluorescentes de la presente invención proporciona un sistema de detección de la actividad PKC. Cuando se fosforila por la PKC, la proteína de fusión de trasloca de la membrana plasmática al citosol. Esta traslocación puede ser rastreada mediante microscopía de fluorescencia convencional o microscopía confocal, por ejemplo, mediante el uso de tecnología Cellomics Inc., u otros sistemas de escrutinio de alto contenido (tales como los de Universal Imaging Corp., o Becton Dickinson). El sistema informador tiene aplicación en el escrutinio de alto contenido para inhibidores de la PKC, y como indicador de la actividad PKC para los análisis de escrutinio de los reactivos potenciales que interfieren con esta ruta de transducción de la señal. Los métodos para utilizar proteínas fluorescentes como biosensores también incluyen los descritos en las Patentes de los Estados Unidos Núms. 5.972.638; 5.824.485 y 5.650.135 (así como las referencias allí citadas) cuyas descripciones se incorporan a la presente memoria como referencia.

Las proteínas fluorescentes de la presente invención también encuentran uso en aplicaciones que implican la selección automatizada de matrices de células que expresan los grupos informadores fluorescentes mediante el uso de imágenes microscópicas y análisis electrónico. Se puede utilizar el escrutinio para el descubrimiento de fármacos y en el campo de la genómica funcional en el que las proteínas sujeto se utilizan como marcadores de células completas para detectar cambios en la reorganización multicelular y la migración, por ejemplo, en la formación de túbulos multicelulares (formación de vasos sanguíneos) por las células endoteliales, la migración de células a través del sistema Fluoroblok Insert (Becton Dickinson Co.), la curación de heridas, o la valoración morfométrica del crecimiento de las neuritas. El escrutinio también puede ser empleado cuando las proteínas de la presente invención se utilizan como marcadores fusionados a péptidos (tal como en secuencias de localización) o proteínas que detectan cambios en la localización intracelular como indicador de la actividad celular, por ejemplo, en la transducción de la señal, tal como la quinasa y la traslocación del factor de transcripción tras los estímulos. Los ejemplos incluyen la proteína quinasa C, la proteína quinasa A, el factor de transcripción NFκB y NFAT; proteínas del ciclo celular, tales como la ciclina A, la ciclina B1 y la ciclina E; escisión por la proteasa con posterior movimiento del sustrato escindido; fosfolípidos, con marcadores para estructuras intracelulares tales como el retículo endoplásmico, el aparato de Golgi, las mitocondrias, los peroxisomas, el núcleo, los nucleolos, la membrana plasmática, las histonas, los endosomas, los lisosomas, o los microtúbulos.

Las proteínas de la presente invención también se pueden utilizar en el escrutinio de alto contenido para detectar la localización simultánea de otras proteínas de fusión fluorescentes con marcadores de localización como indicadores de los movimientos de las proteínas/péptidos intracelulares fluorescentes o solamente como marcadores. Los ejemplos de las aplicaciones que implican el escrutinio automatizado de matrices de células en las que las proteínas fluorescentes sujeto encuentran uso incluyen las Patentes de los Estados Unidos Núm. 5,989,835; así como el documento WO 0017624; el documento WO 00/26408; el documento WO 00/17643; y el documento WO 00/03246; cuyas descripciones se incorporan a la presente memoria como referencia.

Las proteínas fluorescentes de la presente invención también encuentran uso en análisis de escrutinio de alto rendimiento. Las proteínas fluorescentes sujeto son proteínas estables con vidas medias de más de 24 horas. También se proporcionan las versiones desestabilizadas de las proteínas fluorescentes sujeto con vidas medias reducidas que se pueden utilizar como informadores de la transcripción para el descubrimiento de fármacos. Por ejemplo, una proteína de acuerdo con la invención sujeto se puede fusionar con una supuesta secuencia señal proteolítica derivada de una proteína con vida media más corta, como una secuencia PEST del gen de la ornitina descarboxilasa de ratón, una caja de destrucción de ciclina B1 de ratón o ubiquitina, etc. Para una descripción de las proteínas desestabilizadas y los vectores que se pueden emplear para producir las, véanse, por ejemplo, la Patente de los Estados Unidos Núm. 6.130.313; cuya descripción se incorpora en la presente memoria como referencia. Los promotores en las rutas de transducción de la señal se pueden detectar utilizando versiones desestabilizadas de las proteínas fluorescentes sujeto para el escrutinio de fármacos tales como, por ejemplo, AP1, NFAT, NFκB, Smad, STAT, p53, E2F, Rb, myc, CRE, ER, GR y TRE, y similares.

Las proteínas de la presente invención se pueden utilizar como etiquetas fotoactivadas para el fotomarcado *in vivo* preciso y para seguir el tráfico de proteínas, orgánulos o células como describen, por ejemplo, Patterson y Lippincott-Scott, Science, 13:1873-77 (2002) Y Ando, *et al.*, Proc. Natl. Acad. Sci. EE.UU., 99:12651-56 (2002).

Adicionalmente, las proteínas sujeto pueden utilizarse como detectores de segundos mensajeros mediante la fusión de las proteínas sujeto a dominios específicos tales como el dominio de unión PKCgamma Ca, el dominio de unión PKCgamma DAG, el dominio SH2 o el dominio SH3, etc.

- 5 Se pueden preparar formas secretadas de las proteínas sujeto mediante la fusión de las secuencias conductoras secretadas a las proteínas sujeto para construir las formas secretadas de las proteínas sujeto, que a su vez se pueden utilizar en una variedad de aplicaciones diferentes.

10 Las proteínas sujeto también encuentran uso en aplicaciones de clasificación de células activadas por fluorescencia (FACS). En tales aplicaciones, la proteína fluorescente sujeto se utiliza como una etiqueta para marcar una población de células y la población de células etiquetada resultante se clasifica con un dispositivo de clasificación de células activadas por fluorescencia, como se conoce en la técnica. Los métodos FACS se describen en las Patentes de los Estados Unidos Núms. 5.968.738 y 5.804.387; cuyas descripciones se incorporan a la presente memoria como referencia.

15 Las proteínas sujeto también encuentran uso como marcadores *in vivo* en animales transgénicos. Por ejemplo, la expresión de la proteína sujeto puede estar guiada por promotores específicos del tejido, cuando tales métodos encuentran uso en investigaciones de terapia génica, tales como las pruebas de la eficacia de la expresión transgénica, entre otras aplicaciones. Una solicitud representativa de proteínas fluorescentes en animales transgénicos no humanos que ilustra estas aplicaciones se encuentra en el documento WO 00/02997, cuya descripción se incorpora a la presente memoria como referencia.

20 Otras aplicaciones de las proteínas de la presente invención incluyen el uso como marcadores tras la inyección en células o animales y en la calibración para mediciones cuantitativas; como marcadores o informadores en dispositivos biosensores de oxígeno para el control de la viabilidad celular; como marcadores o etiquetas para animales, mascotas, juguetes, alimentos, y similares.

30 Las proteínas fluorescentes sujeto también encuentran uso en análisis de escisión con proteasa. Por ejemplo, se pueden desarrollar análisis de fluorescencia desactivada por escisión utilizando las proteínas sujeto, donde las proteínas sujeto se diseñan para incluir una secuencia de escisión específica de la proteasa, sin destruir el carácter fluorescente de la proteína. Tras la escisión de la proteína fluorescente por una proteasa activada, la fluorescencia se reduciría drásticamente debido a la destrucción del cromóforo funcional. Como alternativa, se puede desarrollar fluorescencia activada por escisión utilizando las proteínas de la presente invención, donde las proteínas están diseñadas para contener una secuencia espaciadora adicional en íntima proximidad/o dentro del cromóforo. La actividad fluorescente de esta variante se redujo significativamente, debido a que porciones del cromóforo funcional son divididas por el separador. El espaciador está enmarcado por dos sitios de escisión específicos de la proteasa idénticos. Después de la escisión a través de la proteasa activada, el espaciador se cortaría y las dos “subunidades” residuales de la proteína fluorescente serían capaces de volver a ensamblarse para generar una proteína fluorescente funcional. Las dos aplicaciones anteriormente mencionadas podrían ser desarrolladas en análisis para una variedad de diferentes tipos de proteasas, como las caspasas y otras.

45 Las proteínas sujeto también se pueden utilizar en análisis para determinar la composición de fosfolípidos en membranas biológicas. Por ejemplo, las proteínas de fusión de las proteínas del sujeto (o cualquier otro tipo de modificación covalente o no covalente de las proteínas sujeto) que permiten la unión a fosfolípidos específicos para localizar/visualizar los patrones de distribución de fosfolípidos en membranas biológicas, al tiempo que permite la localización simultánea de proteínas de membrana en placas de fosfolípidos específicas, se puede lograr con las proteínas sujeto. Por ejemplo, el dominio PH de GRP1 tiene una gran afinidad por fosfatidil-inositol tri-fosfato (PIP3), pero no por PIP2. Como tal, se puede construir una proteína de fusión entre el dominio PH de GRP1 y las proteínas sujeto para etiquetar específicamente áreas ricas en PIP3 en las membranas biológicas.

50 Sin embargo, otra aplicación más de las proteínas sujeto es como reloj fluorescente, en el que el cambio de un color fluorescente a otro (por ejemplo, de verde a rojo), concomitante con el envejecimiento de la proteína fluorescente se utiliza para determinar la activación/desactivación de la expresión génica, tal como la expresión génica evolutiva, la expresión génica dependiente del ciclo celular, la expresión génica específica del ritmo circadiano, y similares.

55 Los anticuerpos de la invención sujeto, descritos anteriormente, también encuentran uso en varias aplicaciones, incluyendo la diferenciación de las proteínas sujeto de otras proteínas fluorescentes.

60 Ejemplo

Se recogieron algunos ejemplares de hidromedusas grandes en la costa Rusa del Mar de Japón cerca de Vladivostok en agosto de 2001. Un conjunto de características permite la identificación de estas medusas, como *Aequorea coerulescens* (Kramp, Dana Rept., 72:201-202 (1968); Pogodin and Yakovlev, Rus. J. Mar. Biol., 25:417-419 (1999)). Aunque *A. coerulescens* y *A. Victoria* (sinónimos de *A. forscalea*, *A. aequorea*) son similares, algunas de sus características son muy diferentes. La diferencia más evidente es el número de tentáculos: *A. Victoria* lleva sólo un tentáculo por canal radial, mientras que *A. coerulescens* posee 4-6 tentáculos entre cada par de canales radiales adyacentes.

Los especímenes capturados de *A. coerulescens* eran bioluminiscentes. En contraste con *A. Victoria*, presentaban una luminiscencia azul en lugar de verde. No se observó fluorescencia detectable en las medusas *A. coerulescens* a la luz ultravioleta o cuando se utilizó un microscopio de fluorescencia. Sin embargo, un anticuerpo monoclonal contra GFP de *A. Victoria* detectó una proteína de tipo GFP en el extracto proteico de *A. coerulescens*. La Figura 11 muestra un análisis de proteínas mediante electroforesis en gel de acGFP. La Figura 11 es un análisis mediante transferencia Western de extracto de proteína soluble de *A. coerulescens* utilizando anticuerpos contra la GFP de *A. Victoria*. La Calle 1 es GFP de *A. Victoria* recombinante purificada, la Calle 2 es el extracto de *A. coerulescens*.

Para clonar la proteína de tipo GFP, se realizó la PCR con cebadores degenerativos correspondientes a las secuencias de aminoácidos conservativas. Se clonó un ADNc que codificaba una proteína de tipo GFP. La secuencia de nucleótidos y aminoácidos de la proteína acGFP de tipo salvaje se muestra en la Figura 1. La proteína acGFP demostró una similitud muy alta con la GFP, con 92% de similitud de la secuencia de aminoácidos; véase la Figura 2. Todos los residuos clave conocidos, incluyendo Ser65, Tyr66 y Gly67 formadores del cromóforo, la Arg96 y el Glu222 invariantes evolutivos, y los residuos espacialmente próximos al cromóforo, His148, Phe165, Ile167 y Thr203 se encontraron sin cambios en acGFP. Sólo tres aminoácidos interiores difirieron entre estas proteínas.

Teniendo en cuenta la muy elevada similitud de secuencia, se esperaba que las propiedades espectrales de acGFP fueran muy similares a las de GFP. No obstante, las colonias *E. coli* que expresan acGFP no mostraron ni fluorescencia ni coloración. La explicación más simple de este hecho -plegamiento de acGFP insatisfactorio en *E. coli*- fue sólo parcialmente correcta, como se demostró en experimentos adicionales.

La presencia de ADNc de acGFP en algunas partes de la medusa se sometió a ensayo utilizando cebadores específicos para PCR. Se sometieron a ensayo tres muestras de ADNc correspondiente al borde de la umbrela, el canal radial, y el disco oral de la medusa. El ADNc de acGFP se detectó claramente en el borde de la umbrela, pero estuvo ausente en las otras dos muestras. Así, la distribución de acGFP dentro de *A. coerulescens* es similar a la distribución de GFP en *A. Victoria*, que forma un anillo fluorescente en el borde de la umbrela de *A. Victoria*.

La mutagénesis aleatoria de acGFP produjo muchos clones fluorescentes verdes, algunos de los cuales se caracterizaron. Sus propiedades y posibles aplicaciones son similares a los del mutante de GFP mejorado (EGFP) de *A. Victoria*.

El mutante Z1 contenía una sustitución de un aminoácido, E222G (una glicina por un ácido glutámico en la posición 222), como se observa en la Figura 3. La proteína Z1 mutante tenía poca luminosidad, un plegamiento muy lento y requirió una temperatura de menos de 20°C para su maduración. Después de crecer a 30°C, las colonias de *E. coli* que expresan Z1 se deben almacenar durante 3-5 días a temperatura ambiente o a 4°C para que la fluorescencia se haga visible. Los espectros de excitación y de emisión para el mutante Z1 tienen máximos a 480 y 504 nm, respectivamente (véase la Figura 4).

El mutante Z2 contenía dos sustituciones de aminoácidos, específicamente, N19D (un ácido aspártico por una asparagina en la posición 19) y E222G (una glicina por un ácido glutámico en la posición 222) como se observa en la Figura 5. El mutante Z2 poseía una baja luminosidad, eficiencia de plegamiento muy lento y requiere una temperatura inferior a 20°C durante la maduración. Después de crecer durante la noche a 37°C, las colonias de *E. coli* que expresan Z2 se deben almacenar durante 3-5 días a temperatura ambiente o a 4°C para que la fluorescencia se haga visible. Los espectros de excitación y de emisión para el mutante Z2 son muy similares a los del mutante Z1.

El mutante G1 tiene sustituciones V11I (una isoleucina por una valina en la posición 11), K101E (un ácido glutámico de una lisina en la posición 101) y E222G (una glicina por un ácido glutámico en la posición 222) como se observa en la Figura 6. El mutante G1 se generó en una ronda independiente de mutagénesis al azar. Puesto que se encontró la sustitución E222G en tres de los mutantes, parece probable que esta mutación fuera importante para la fluorescencia verde de estos mutantes. El mutante G1 posee una luminosidad bastante baja. La fluorescencia de este mutante se hace visible el primer día después del crecimiento durante la noche de *E. coli*. Los espectros de excitación y de emisión de los mutantes G1 son muy similares a los del mutante Z1.

El mutante G2 tiene sustituciones V11I (una isoleucina por una valina en la posición 11), F64L (una leucina por fenilalanina en la posición 6A), K101E (un ácido glutámico por lisina en la posición 101) y E222G (una glicina por un ácido glutámico en la posición 222) como se observa en la Figura 7. El mutante G2 se generó basándose en el mutante G1 con una segunda ronda de mutagénesis al azar. En comparación con G1, la proteína mutante G2 posee características de luminosidad y de velocidad de plegamiento de la proteína mejoradas. Curiosamente, G2 contiene la sustitución F64L que es también característica de GFP mejorada (EGFP) (Cormack *et al.*, Gene, 173:33-38 (1996); Yang *et al.*, Nuc. Acids Res., 24:4592-4593 (1996)). Las propiedades espectrales de G2 se muestran en la Tabla 1, más abajo. Los espectros de excitación-emisión de G2 se muestran en la Figura 8.

El mutante G22 tiene sustituciones V11I (una isoleucina por una valina en la posición 11), F64L (una leucina por una fenilalanina en la posición 64), K101E (un ácido glutámico por una lisina en la posición 101), T206A (una alanina por una treonina en la posición 206), y E222G (una glicina por un ácido glutámico en la posición 222) como se observa en la Figura 9. El mutante G22 se generó a partir del mutante G2 con una tercera ronda de mutagénesis al azar. En comparación con G2, G22 posee una luminosidad mejorada aún mayor. Las propiedades espectrales de G22 se presentan en la Tabla 1. Los espectros de excitación-emisión para G22 se muestran en la Figura 10. Un coeficiente

de extinción de $50.000 \text{ M}^{-1}\text{cm}^{-1}$ y un rendimiento cuántico de 0,55 hacen esta proteína casi tan brillante como la GFP mejorada ampliamente utilizada. Los ensayos de filtración en gel, así como la SDS-PAGE de la proteína G22 mutante no calentada demostraron que G22 es monomérica. En la Figura 11B se compara la movilidad de muestras de la proteína calentada (calles 1-3) frente a no calentadas (calles 4-6). Las calles 1 y 4 son GFP de *A. Victoria*, las calles 2 y 5 son el mutante G22, y las calles 3 y 6 son el mutante G22-G222E. La tinción con azul de Coomassie se muestra a la izquierda y la fluorescencia de las proteínas no calentadas bajo luz ultravioleta se muestra a la derecha.

Todos los mutantes fluorescentes de acGFP mencionados anteriormente tienen espectros de excitación y emisión similares, alcanzando un máximo de 470-480 nm y 500-510 nm, respectivamente. La forma de sus espectros de excitación son similares a los de la GFP mejorada, pero no a los de GFP de *A. Victoria* de tipo salvaje. Es probable que el fluoróforo de los mutantes acGFP esté siempre en forma desprotonada, como se ha demostrado que está para GFP mejorada. Una posible explicación es la ausencia de Glu222, lo que probablemente puede ser importante para la transferencia de protones (Ehrig *et al.*, FEBS Lett., 367:163-166 (1995)).

Para aclarar la importancia de la sustitución de E222G para la fluorescencia, se realizó una sustitución G222E inversa para G22 mutante (sustituciones V11I, F64L, K101E, T206A). El ácido nucleico y las secuencias de aminoácidos de este mutante inverso G22-E222G se muestran en la Figura 12. La mutación inversa transforma fácilmente la proteína mutante G22-E222G a un estado incoloro. Las colonias de *E. coli* que expresan G22-G222E no presentan ni coloración ni fluorescencia detectable.

G22-G222E se expresó y se plegó en *E. coli* a 37°C sin problema, como lo demuestra el alto rendimiento de la proteína recombinante soluble (alrededor del 20% de proteínas totales). G22-G222E purificada muestra un espectro de absorción con un pico principal a 280 nm y un pico secundario a 390 nm (véase la Figura 13A). La proteína G22-G222E desnaturizada con álcali mostró un pico de absorción a 446 nm, que aparentemente corresponde a la forma aniónica del cromóforo de GFP. Suponiendo un coeficiente de extinción $44.000 \text{ M}^{-1}\text{cm}^{-1}$ para el cromóforo, se estimó que la extinción de G22-G222E nativa a 390 nm era de $33.000 \text{ M}^{-1}\text{cm}^{-1}$. La relación observada entre los picos a 280 nm y 390 nm (se calculó que el coeficiente de extinción molar a 280 nm era de $23.500 \text{ M}^{-1}\text{cm}^{-1}$) mostrando que sólo aproximadamente el 3% del G22-G222E soluble existía en una forma madura. La excitación de 390 nm condujo a un doble pico de fluorescencia débil con color a 460 nm y 505 nm con un rendimiento cuántico de 0,07 (véase la Figura 13B).

Es bien sabido que las proteínas de tipo GFP conservan sus propiedades espectrales y el estado de oligomerización en condiciones de SDS-PAGE comunes siempre que las muestras de proteínas no sean calentadas antes de cargarlas en el gel. Esta prueba se utiliza para examinar el estado de plegamiento de G22-G222E. La electroforesis en gel demostró una clara diferencia en la movilidad de proteínas no desnaturizadas y desnaturizadas (véase la Figura 11). Además, la banda de la proteína G22-G222E no calentada produjo una fluorescencia muy débil bajo luz ultravioleta (de nuevo, véase la Figura 11). Aproximadamente 97% de esta proteína está presente en una forma no absorbente, estos resultados indican que la conformación de esta forma no absorbente está próxima al estado nativo, pero no al estado desnaturizado.

Alentados por estos resultados, se repitieron los intentos de obtener y caracterizar la acGFP de tipo salvaje recombinante. El crecimiento de *E. coli* que expresa acGFP a temperatura ambiente sin inducción, seguido de una incubación de varios días a 4°C, dio como resultado la aparición de una pequeña fracción de acGFP soluble (alrededor del 5% de acGFP total). Las formas de los espectros de absorción y fluorescencia para acGFP de tipo salvaje soluble son muy similares a los de G22-G222E. Se concluyó que el mutante G22-G222E refleja las propiedades de la acGFP naturales pero posee plegamiento de proteínas y estabilidad a la temperatura mejorados cuando se expresa en *E. coli*.

Estos datos demostraron que G22-G222E soluble así como acGFP de tipo salvaje existe en dos formas. La mayoría de estas proteínas están presentes en una forma plegada pero inmadura sin un cromóforo detectable espectralmente. La forma que absorbe a 390 nm minoritaria contiene un cromóforo de tipo GFP en un estado neutro y posee doble fluorescencia con color débil.

Un nuevo tipo de fotoconversión se observó en G22-G222E mutante. La irradiación de una muestra de proteína G22-G222E con 250-300 nm de luz UV dio como resultado la aparición de un pico de 480 nm en los espectros de absorción/excitación. Obsérvese en la Figura 14 el espectro de excitación de G22-G222E antes de la irradiación (línea 1) y el cambio gradual de la curva debido a la irradiación de la muestra de proteína con luz de 250-300 nm. La línea sin número representa el espectro de emisión después de la fotoconversión (excitación a 480 nm). Esto puede proceder de una forma espectralmente no detectable, inmadura de la proteína, puesto que el pico de absorción a 390 nm no disminuyó durante esta fotoconversión. La excitación en el pico de 480 nm produjo fluorescencia verde a 505 nm con un alto rendimiento cuántico (0,45). Se logró un aumento de la intensidad de la fluorescencia verde inducido por UV de más de 1000 veces (excitación a 480 nm).

El mutante G22-G222E/Y22 0L tiene sustituciones V11I (una isoleucina por una valina en la posición 11), F64L (una leucina por fenilalanina en la posición 64), Vg8A (una alanina por una valina en la posición 68) K1 01 E (un ácido glutámico por una lisina en la posición 101), T206A (una alanina por una treonina en la posición 206), y Y220L (una leucina por una tirosina en la posición 220) en comparación con acGFP de tipo salvaje, como se observa en la Figura 15. Este mutante demostraba un plegamiento de proteína a 37°C y poseía una clara fluorescencia de color verde a 508 nm. El espectro de excitación para G22-G222E/Y220L tenía un pico principal a 396 nm y un pico secundario

a 493 nm (10:1) (véase la Figura 16, donde los espectros de excitación son la línea discontinua y los espectros de emisión son la línea continua). Las propiedades espectrales de este mutante fueron muy similares a la de GFP de tipo salvaje de *Aequorea victoria*.

5 Usando el mutante G22-G222E/Y22 0L como base, se obtuvo el mutante 220-II-5 que tenía las sustituciones V11I (una isoleucina por una valina en la posición 11), F64L (una leucina por una fenilalanina en la posición 64), K101E (un ácido glutámico por una lisina en la posición 101), E115K (una lisina por un ácido glutámico en la posición 115), H148Q (una glutamina por una histidina en la posición 148), T206A (una alanina por una treonina en el puesto 206), Y220L (una leucina por una tirosina en la posición 220), F221L (una leucina por una fenilalanina en la posición 221), y K238Q (una glutamina por una lisina en la posición 238). Las secuencias de ácidos nucleicos y de aminoácidos de este mutante están en la Figura 17. Este mutante tiene un pico de excitación principal a 395 nm, y no hay ningún pico de excitación cerca de 480 nm, con el pico de emisión a 512 nm (véase la Figura 18A). Es probable que la supresión del pico de excitación con la mayor longitud de onda pueda ser explicada por la sustitución H148Q que da como resultado la desaparición de la fracción del cromóforo cargado. Después de varios minutos de la irradiación relativamente intensa con luz de aproximadamente 400 nm bajo un microscopio de fluorescencia, el espectro de excitación de G22-G222E/Y220L mutante cambió. Hubo una disminución simultánea del pico a 395 nm y la aparición del pico de excitación a 480 nm (véase la Figura 18B). Como resultado, se pudo obtener un contraste de más de 100 veces en la luminosidad de la fluorescencia de la emisión a 510 nm en la luz de excitación de 480 nm, en comparación con antes y después de la irradiación de la luz intensa de 400 nm de longitud de onda. La Figura 25 muestra dos colonias de *E. coli* que expresan 220-II-5 mutante bajo un microscopio de fluorescencia. Las dos áreas en la colonia superior fueron fotoactivadas preliminarmente por una luz intensa de 400 nm. Como tales, los mutantes G22-G22E/Y2201 se pueden utilizar como marcador fluorescente fotoactivado para el fotomarcaje de organismos vivos de una manera similar a los métodos publicados recientemente para el uso del mutante PA-GFP por Patterson and Lippincott-Schwartz, Science, 13:1873-1877 (2002).

25 El mutante de proteína fluorescente cian CFP-rand3 tiene sustituciones en V11I (isoleucina por valina en la posición 11), T62A (alanina por treonina en la posición 62), F64L (leucina por fenilalanina en la posición 64), K101 E (ácido glutámico por lisina en la posición 101), N121S (serina de asparagina en la posición 121), H148T (treonina por histidina en la posición 148), E172K (lisina por ácido glutámico en la posición 172) y T206A (alanina por treonina en la posición 206). Las secuencias de aminoácidos y ácidos nucleicos de CFP-rand3 mutante se muestran en la Figura 19. Este mutante tiene un pico de excitación a 402 nm, con un pico de emisión sencillo a 467 nm (véase la Figura 20).

35 Se generó otra proteína fluorescente cian mutante, CFP-3, que tenía las sustituciones V11I (isoleucina por valina en la posición 11), F64L (una leucina por una fenilalanina en la posición 64), K101E (ácido glutámico por lisina en la posición 101), H148S (serina por histidina en la posición 148), F165L (leucina por fenilalanina en la posición 165), E172A (alanina por ácido glutámico en la posición 172), y T206A (alanina por treonina en la posición 206). Las secuencias de aminoácidos y ácidos nucleicos para este mutante se muestran en la Figura 21. Este mutante tiene un pico de excitación a 390 nm, con un pico de emisión sencillo a 470 nm, Figura 22A. Después de varios minutos de irradiación con luz relativamente intensa a 400 nm, se observaron los siguientes cambios en los espectros de excitación y emisión: (i) disminución considerable del pico de excitación a 390 nm, y (ii) aparición de un pico de excitación a 480 nm con emisiones a 505 nm (véase la Figura 22B).

45 Como resultado, se puede obtener un contraste de más de 30 veces en la luminosidad de la fluorescencia de la emisión a 505 nm utilizando luz de excitación de 480 nm, en comparación con los espectros antes y después de la irradiación intensa de 405 nm. El cambio simultáneo de los parámetros de excitación y emisión transforma el mutante cian CFP-3 en una proteína fluorescente verde en respuesta a la irradiación intensa de 400 nm. Por lo tanto, el mutante CFP-3 se puede utilizar como marcador fluorescente fotoactivado o "foto-conmutado" para el fotomarcaje de organismos vivos.

50 En un experimento ilustrativo, el mutante G22 se utilizó como una etiqueta fluorescente para someter a ensayo la expresión de las proteínas en células de mamíferos. Inesperadamente, sin embargo, G22 mutante produjo una señal fluorescente muy baja en las líneas celulares humanas. Esto probablemente puede ser explicado por el uso no óptimo de los codones o por la presencia de un intrón críptico en el gen mutante G22. Para superar estos dos problemas, se sintetizó un gen mutante G22-h incorporando el uso de codón optimizado de mamífero. Las secuencias de aminoácidos y de ácidos nucleicos para este mutante humanizado se muestran en la Figura 23. La expresión transitoria de G22-H en diferentes líneas celulares mostró una señal de color verde brillante sin agregación.

60 La Figura 24 muestra la expresión transitoria del mutante G22-H (paneles AE), y una proteína de fusión G22- β -actina (F), en diferentes líneas celulares de mamífero. Panel A-293T, panel B-vero; panel C-3T3; panel D-L929, panel E-cos1; panel F-3T3. La fluorescencia fue claramente detectable 24 horas después de la transfección. No se observó toxicidad. La capacidad de G22-h para etiquetar proteínas se demostró mediante la construcción de una proteína de fusión con β -actina citoplasmática. La expresión transitoria de esta fusión en células 3T3 mostró un marcaje de la actina luminoso y preciso (véase la Figura 24F). Las fibras de estrés, los contactos focales y los procesos celulares fueron claramente visibles. No se observó detrimento en la adherencia o la vitalidad celulares, ni se observó ninguna agregación proteica no específica.

Métodos

El ARN total se aisló utilizando un kit NucleoSpin RNA II (Clontech) a partir de una pequeña vivisección de un organismo *Aequorea coerulescens* que incluyó el borde de la umbrela y el canal radial. El ADNc se sintetizó y se amplificó con un kit SMART PCR cDNA Synthesis (Clontech). Se obtuvo un fragmento del gen de la proteína fluorescente novedoso mediante PCR con cebadores degenerados. Se utilizó un método PCR RACE por etapas para clonar el fragmento del extremo 5' del ADNc diana. La secuencia de nucleótidos del ADNc que codifica la proteína fluorescente novedosa, acGFP, se ha presentado a GenBank con el número de acceso AY151052. Para la expresión bacteriana de acGFP, se amplificó la región codificante completa utilizando cebadores específicos y se clonó en el vector pQE30 (Qiagen).

Se utilizaron varios kits PCR Randomm Mutagenesis (Clontech) para la mutagénesis aleatoria de acGFP, en condiciones óptimas para 5-6 mutaciones por 1000 pares de bases. Se examinaron visualmente colonias de *E. coli* que expresaban proteínas mutantes con un estereomicroscopio fluorescentes SZX-12 (Olympus). Se seleccionaron las variantes más luminosas y se sometieron a una ronda adicional de mutagénesis al azar. Se realizó la mutagénesis dirigida al sitio mediante PCR utilizando el método de extensión del solapamiento, con cebadores que contenían las sustituciones diana apropiadas (véase, por ejemplo, Ho *et al.*, *Gene*, 77:51-59 (1989).

Las proteínas fusionadas a una etiqueta de seis histidinas N-terminal se expresaron en la Cepa XL1 blue de *E. coli* (Invitrogen) y se purificaron utilizando una resina de afinidad con metal TALON (Clontech). Los espectros de absorción fueron registrados con un Beckman DU520 UV/VIS. Se utilizó un espectrofotómetro Varian Cary Eclipse Fluorescence para medir los espectros de excitación-emisión.

Para la determinación del coeficiente de extinción molar, se utilizó la estimación de la concentración de cromóforo maduro. Las proteínas se desnaturalizaron con álcali con un volumen igual de NaOH 2 M. En estas condiciones, el cromóforo de GFP de *A. victoria* absorbe a 446 nm y su coeficiente de extinción molar es igual a $44.000 \text{ M}^{-1}\text{cm}^{-1}$ (Ward *et al.*, *Photochem. Photobiol.*, 31:611-615 (1980)). Se midieron los espectros de absorción para las proteínas nativas y desnaturalizadas con álcali. Los coeficientes de extinción molar para el estado nativo se estimaron mediante la absorción de proteínas desnaturalizadas como base. Para la determinación del rendimiento cuántico, se comparó la fluorescencia de los mutantes con la de GFP mejorada (rendimiento cuántico 0,60).

Se realizó la fotoconversión inducida por UV de acGFP-G222E utilizando un espectrofotómetro Cary Eclipse Fluorescence. La muestra de la proteína se irradió durante varias horas con luz de longitud de onda de 250-300 nm en el modo barrido (rendija de excitación de 20 nm, velocidad de barrido de 30 nm/min, con un promedio de tiempo de 1 segundo, en modo ciclo).

Las muestras de proteína purificada (~1 mg/ml) se cargaron en una columna Sephadex-100 (0,7 x 60 cm) y se hicieron eluir con tampón de fosfato 50 mM (pH 7,0) con NaCl 100 mM. EGFP, HcRed1, y DsRed2 (Clontech) se utilizaron como patrones para el monómero, dímero y tetramero, respectivamente.

Para el análisis de proteínas en gel, las muestras calentadas y sin calentar se cargaron en una SDS-PAGE 12% común, y la electroforesis se llevó a cabo a 15 mA/gel. Para la transferencia Western, las proteínas se transfirieron a una membrana Hybond C (Amersham), utilizando procedimientos convencionales. Las membranas se sondearon con anticuerpos de ratón (Clontech) frente a GFP (1:2500), y luego con anticuerpos anti-ratón conjugados con HRP (Amersham) a 1:2500. Para desarrollar el patrón de tinción, se utilizó un sistema de análisis por transferencia Western ECL (Amersham Pharmacia Biotech), incluyendo los reactivos de detección 1, 2 e Hyperfilm ECL.

Para su expresión en células eucariotas, se clonó acGFP en vectores pEGFP-C1 y pEGFP-Actin (CLONTECH) entre los sitios de restricción *AgeI* y *BglIII* (en lugar de la región codificante de EGFP). Se utilizaron las siguientes líneas celulares: células epiteliales renales humanas 293T, fibroblastos de embrión de ratón 3T3, fibroblastos subcutáneos de múrido L929, células epiteliales de riñón de mono verde africano Vero, y fibroblastos de riñón de mono verde africano COS1. Las células se transfectaron con reactivo LipofectAMINE (Invitrogen) y se evaluaron 20 horas después de la transfección. Se utilizó un microscopio Olympus CK40 fluorescente equipado con una cámara CCD DP-50 (Olympus) para obtener imágenes de la célula.

Tabla 1. Propiedades espectrales de los mutantes acGFP en comparación con EGFP.

Especie	Nombre de la proteína	Máximo de absorción, nm	Máximo de emisión, nm	Coefficiente de extinción máximo, M ⁻¹ cm ⁻¹	Rendimiento cuántico	Luminosidad relativa **
<i>Aequorea victoria</i>	EGFP *	488	509	53.000	0,60	1
<i>Aequorea</i>	De tipo salvaje	390	460, 505	nt	nt	nt
<i>coerulescens</i>	acGFP					
	Mutante G2	475	504	58.000	0,38	0,69
	Mutante G22	480	505	50.000	0,55	0,86
	Mutante	390	460, 505	33.000	0,07	0,07
	G22-G222E***					
	Mutante	402	467	35.000	0,30	0,33
	CFP-rand3					
* Datos de referencia: Patterson, G., Day, R.N., and Piston, D. (2001) Espectros de la proteína fluorescente. J. Cell. Sci. 114, 837-838						
** En comparación con la luminosidad (coeficiente de extinción multiplicado por el rendimiento cuántico) de EGFP.						
*** Datos sobre la fracción madura del mutante (aproximadamente 3% de la proteína total)						

REIVINDICACIONES

1. Un ácido nucleico aislado que codifica un péptido de *Aequorea coerulescens*, seleccionado entre:

- (a) un ácido nucleico que codifica una proteína que comprende la secuencia de aminoácidos mostrada en la Figura 1, y
- (b) un ácido nucleico que comprende la secuencia de nucleótidos mostrada en la Figura 1.

2. Un ácido nucleico aislado que comprende una secuencia de ácido nucleico que codifica un mutante fluorescente de una proteína de *Aequorea coerulescens* con la secuencia de aminoácidos mostrada en la Figura 1 seleccionado del grupo consistente en:

- (a) un ácido nucleico que codifica una proteína fluorescente que tiene una identidad de secuencia de al menos 95% con una secuencia de aminoácidos seleccionada entre las secuencias de aminoácidos mostradas en las Figuras 3, 5, 6, 7, 9, 12, 15, 17, 19, o 21, o
- (b) un ácido nucleico que comprende la secuencia de nucleótidos seleccionada entre las secuencias de nucleótidos mostradas en las Figuras 3, 5, 6, 7, 9, 12, 15, 17, 19, 21, o 23.

3. Un ácido nucleico aislado de la reivindicación 2, donde dicho ácido nucleico codifica una secuencia de aminoácidos seleccionada del grupo que consiste en las secuencias de aminoácidos mostradas en las Figuras 3, 5, 6, 7, 9, 12, 15, 17, 19 o 21.

4. Una casete de expresión que comprende

- (a) el ácido nucleico de la reivindicación 2; y
- (b) los elementos reguladores necesarios para la expresión del ácido nucleico en la célula.

5. Una casete de expresión que comprende

- (a) el ácido nucleico de la reivindicación 3, y
- (b) los elementos necesarios para la expresión del ácido nucleico en la célula.

6. Una célula, o su progenie, que comprende la casete de expresión de la reivindicación 4.

7. Una célula, o su progenie, que comprende la casete de expresión de la reivindicación 5.

8. Un péptido aislado codificado por el ácido nucleico de la reivindicación 2.

9. Un anticuerpo que se une específicamente al péptido de la reivindicación 8.

10. Un péptido de fusión que incorpora el péptido de la reivindicación 8.

11. Un organismo transgénico no humano capaz de expresar el péptido de la reivindicación 8.

12. Un método para marcar o detectar una molécula biológica que comprende acoplar dicha molécula biológica al péptido de la reivindicación 8.

13. Un método para marcar o detectar una célula o un orgánulo celular que comprende la producción del péptido de la reivindicación 8 en la célula.

14. Un método detectar la expresión génica que comprende la producción del péptido de la reivindicación 8 en la célula.

ES 2 340 154 T3

```

      10      20      30      40      50      60
5' ATTCAAAACACTGCAGAATTTTGGATAGATTTTCCTGCTACTTCACACGCATAAAAGACA

      70      80      90     100     110     120
AGAAAGATGAGTAAAGGAGCAGAACTTTTCCTGAGTTGTCCCAATTCTTATTGAATTA
  M  S  K  G  A  E  L  F  T  G  V  V  P  I  L  I  E  L

      130     140     150     160     170     180
AATGGTGATGTTAATGGGCACAAATTCCTGTCAGTGGAGAGGGCGAAGGTGATGCGACA
  N  G  D  V  N  G  H  K  F  S  V  S  G  E  G  E  G  D  A  T

      190     200     210     220     230     240
TACGGAAAGTTAACCTTAAATTTATTTGCACTACAGGAAACTACCTGTTCCATGGCCA
  Y  G  K  L  T  L  K  F  I  C  T  T  G  K  L  P  V  P  W  P

      250     260     270     280     290     300
ACACTTGTCACTACTTTCTCTTATGGTGTTCAATGCTTTTCAAGATATCCAGATCATATG
  T  L  V  T  T  F  S  Y  G  V  Q  C  F  S  R  Y  P  D  H  M

      310     320     330     340     350     360
AAACAGCATGACTTCTTCAAGAGTGCCATGCCTGAAGGTTATATACAGGAAAGAACTATA
  K  Q  H  D  F  F  K  S  A  M  P  E  G  Y  I  Q  E  R  T  I

      370     380     390     400     410     420
TTTTTCAAAGATGACGGGAACACAGTTCGCGTGCTGAAGTCAAGTTTCAAGGTGATACC
  F  F  K  D  D  G  N  Y  K  S  R  A  E  V  K  F  E  G  D  T

      430     440     450     460     470     480
CTGGTTAATAGAATTGAGTTAACAGGTACTGATTTTAAAGAAGATGGAACATCCTTGGGA
  L  V  N  R  I  E  L  T  G  T  D  F  K  E  D  G  N  I  L  G

      490     500     510     520     530     540
AATAAAATGGAATACAACATAACGCACATAATGTATACATCATGACAGACAAAGCAAAA
  N  K  M  E  Y  N  Y  N  A  H  N  V  Y  I  M  T  D  K  A  K

      550     560     570     580     590     600
AATGGAATCAAAGTTAACTTCAAAATTAGACACAACATTGAAGATGGAAGCGTTCAACTT
  N  G  I  K  V  N  F  K  I  R  H  N  I  E  D  G  S  V  Q  L

      610     620     630     640     650     660
GCAGACCATTATCAACAAAATACTCCAATTGGCGATGGCCCTGTCCTTTTACCAGATAAC
  A  D  H  Y  Q  Q  N  T  P  I  G  D  G  P  V  L  L  P  D  N

      670     680     690     700     710     720
CATTACCTGTCCACACAATCTACCCCTTCCAAAGATCCCAACGAAAAGAGAGATCACATG
  H  Y  L  S  T  Q  S  T  L  S  K  D  P  N  E  K  R  D  H  M

      730     740     750     760     770     780
ATCTATTTTGTAGTTTGTAAACAGCTGCTGCGATTACACATGGCATGGATGAATTATACAAA
  I  Y  F  E  F  V  T  A  A  A  I  T  H  G  M  D  E  L  Y  K

      790     800     810     820     830     840
TAAATGTATAGACTTCAAGTTGACACTAACGTGTCCGAACAATTACTAAAATCTCAGGGT

      850     860     870     880     890     900
TCCTGGTTAAAATCAGGCTGAGATATTATTTACATATTATAGATTATTAGAAATTATTTA

      910     920     930     940
AATACTTTATAGATGTTATTGATAGGGGTTATTTTCTTATT 3'

```

FIG. 1



FIG. 2

ES 2 340 154 T3

1	ATG AGT AAA GGA GCA GAA CTT TTC ACT GGA GCT GTC CCA ATT CTT ATT	48
1	M S K G A E L F T G A V P I L I	16
49	GAA TTA AAT GGT GAT GTT AAT GGG CAC AAA TTC TCT GTC AGT GGA GAG	96
17	E L N G D V N G H K F S V S G E	32
97	GGC GAA GGT GAT GCG ACA TAC GGA AAG TTA ACC CTT AAA TTT ATT TGC	144
33	G E G D A T Y G K L T L K F I C	48
145	ACT ACA GGA AAA CTA CCT GTT CCA TGG CCA ACA CTT GTC ACT ACT TTC	192
49	T T G K L P V P W P T L V T T F	64
193	TCT TAT GGT GTT CAA TGC TTT TCA AGA TAT CCA GAT CAT ATG AAA CAG	240
65	S Y G V Q C F S R Y P D H M K Q	80
241	CAT GAC TTC TTC AAG AGT GGC ATG CCT GAA GGT TAT ATA CAG GAA AGA	288
81	H D F F K S A M P E G Y I Q E R	96
289	ACT ATA TTT TTC AAA GAT GAC GGG AAC TAC AAG TCG CGT GCT GAA GTC	336
97	T I F F K D D G N Y K S R A E V	112
337	AAG TTC GAA GGT GAT ACC CTG GTT AAT AGA ATT GAG TTA ACA GGT ACT	384
113	K F E G D T L V N R I E L T G T	128
385	GAT TTT AAA GAA GAT GGA AAC ATC CTT GGA AAT AAA ATG GAA TAC AAC	432
129	D F K E D G N I L G N K M E Y N	144
433	TAT AAC GCA CAT AAT GTA TAC ATC ATG ACA GAC AAA GCA AAA AAT GGA	480
145	Y N A H N V Y I M T D K A K N G	160
481	ATC AAA GTT AAC TTC AAA ATT AGA CAC AAC ATT GAA GAT GGA ACC GTT	528
161	I K V N F K I R H N I E D G S V	176
529	CAA CTT GCA GAC CAT TAT CAA CAA AAT ACT CCA ATT GGC GAT GGC CCT	576
177	Q L A D H Y Q Q N T P I G D G P	192
577	GTC CTT TTA CCA GAT AAC CAT TAC CIG TCC ACA CAA TCT ACC CTT TCC	624
193	V L L P D N H Y L S T Q S T L S	208
625	AAA GAT CCC AAC GAA AAG AGA GAT CAC ATG ATC TAT TTT GGG TTT GTA	672
209	K D P N E K R D H M I Y F G F V	224
673	ACA GCT GCT GCG ATT ACA CAT GGC ATG GAT GAA TTA TAC AAA TTA	717
225	T A A A I T H G M D E L Y K *	239

FIG. 3

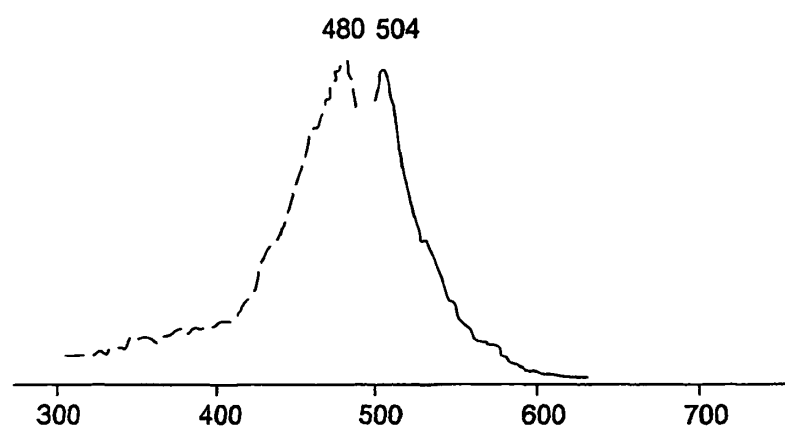


FIG. 4

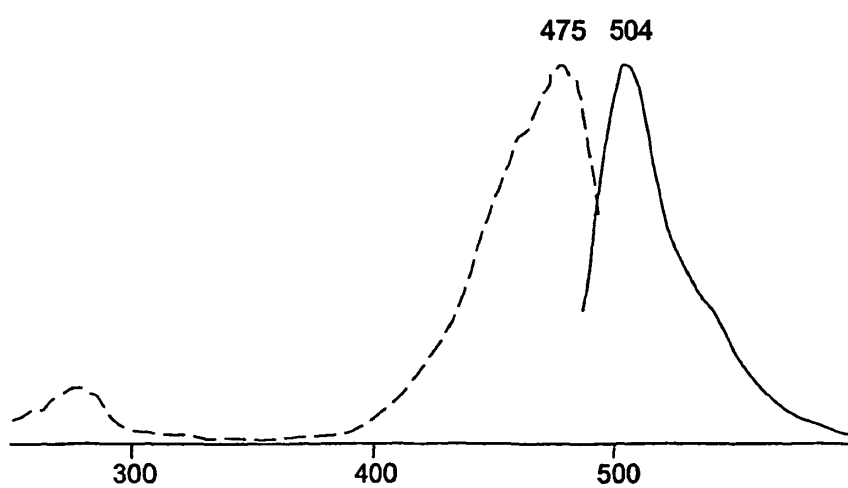


FIG. 8

ES 2 340 154 T3

1	ATG AGT AAA GGA GCA GAA CTT TTC ACT GGA GCT GTC CCA ATT CTT ATT	48
1	M S K G A E L F T G A V P I L I	16
49	GAA TTA GAT GGT GAT GTT AAT GGG CAC AAA TTC TCT GTC AGT GGA GAG	96
17	E L D G D V N G H K F S V S G E	32
97	GGC GAA GGT GAT GCG ACA TAC GGA AAG TTA ACC CTT AAA TTT ATT TGC	144
33	G E G D A T Y G K L T L K F I C	48
145	ACT ACA GGA AAA CTA OCT GTT CCA TGG CCA ACA CTT GTC ACT ACT TTC	192
49	T T G K L P V P W P T L V T T F	64
193	TCT TAT GGT GTT CAA TGC TTT TCA AGA TAT CCA GAT CAT ATG AAA CAG	240
65	S Y G V Q C F S R Y P D H M K Q	80
241	CAT GAC TTC TTC AAG AGT GCC ATG OCT GAA GGT TAT ATA CAG GAA AGA	288
81	H D F F K S A M P E G Y I Q E R	96
289	ACT ATA TTT TTC AAA GAT GAC GGG AAC TAC AAG TCG CGT GCT GAA GTC	336
97	T I F F K D D G N Y K S R A E V	112
337	AAG TTC GAA GGT GAT ACC CTG GTT AAT AGA ATT CAG TTA ACA GGT ACT	384
113	K F E G D T L V N R I E L T G T	128
385	GAT TTT AAA GAA GAT GGA AAC ATC CTT GGA AAT AAA ATG GAA TAC AAC	432
129	D F K E D G N I L G N K M E Y N	144
433	TAT AAC GCA CAT AAT GTA TAC ATC ATG ACA GAC AAA GCA AAA AAT GCA	480
145	Y N A H N V Y I M T D K A K N G	160
481	ATC AAA GTT AAC TTC AAA ATT AGA CAC AAC ATT GAA GAT GCA AGC GTT	528
161	I K V N F K I R H N I E D G S V	176
529	CAA CTT GCA GAC CAT TAT CAA CAA AAT ACT CCA ATT GGC GAT GGC OCT	576
177	Q L A D H Y Q Q N T P I G D G P	192
577	GTC CTT TTA CCA GAT AAC CAT TAC CTG TCC ACA CAA TCT ACC CTT TCC	624
193	V L L P D N H Y L S T Q S T L S	208
625	AAA GAT CCC AAC GAA AAG AGA GAT CAC ATG ATC TAT TTT GGG TTT GTA	672
209	K D P N E K R D H M I Y F G F V	224
673	ACA GCT GCT GCG ATT ACA CAT GGC ATG GAT GAA TTA TAC AAA TAA	717
225	T A A A I T H G M D E L Y K *	239

FIG. 5

1	ATG AGT AAA GGA GCA GAA CTT TTC ACT GGA	48
1	M S K G A E L F T G	16
49	GAA TTA AAT GGT GAT GGT AAT GGG CAC AAA TTC TCT GTC AGT GGA GAG	96
17	E L N G D V N G H K F S V S G E	32
97	GGC GAA GGT GAT GGG ACA TAC GGA AAG TTA ACC CTT AAA TTT ATT TGC	144
33	G E G D A T Y G K L T L K F I C	48
145	ACT ACA GGA AAA CTA TGA CCT GGT CCA TGG CCA ACA CTT GTC ACT ACT	192
49	T T G K L * P V P W P T L V T T	64
193	TTC TCT TAT GGT GGT CAA TGC TTT TCA AGA TAT CCA GAT CHT AUG AAA	240
65	F S Y G V Q C F S R Y P D H M K	80
241	CAG CHT GAC TTC TTC AAG AGT GCC AUG CCT GAA GGT TAT ATA CAG GAA	288
81	Q H D F F K S A M P E G Y I Q E	96
289	AGA ACT ATA TTT TTC GAA GAT GAC GGG AAC TAC AAG TCG CGT GCT GAA	336
97	R T I F F D D G N Y K S R A E	112
337	GTC AAG TTC GAA GGT GAT ACC CIG GGT AAT AGA ATT GAG TTA ACA GGT	384
113	V K F E G D T L V N R I E L T G	128
385	ACT GAT TTT AAA GAA GAT GGA AAC ATC CTT GGA AAT AAA AUG GAA TAC	432
129	T D F K E D G N I L G N K M E Y	144
433	AAC TAT AAC GCA CHT AAT GGA TAC ATC AUG ACA GAC AAA GCA AAA AAT	480
145	N Y N A H N V Y I M T D K A K N	160
481	GGA ATC AAA GTT AAC TTC AAA ATT ACA CAC AAC ATT GAA GAT GGA AGC	528
161	G I K V N F K I R H N I E D G S	176
529	GGT CAA CTT GCA GAC CHT TAT CAA CAA AAT ACT CCA ATT GGC GAT GGC	576
177	V Q L A D H Y Q Q N T P I G D G	192
577	CCT GTC CTT TTA CCA GAT AAC CHT TAC CIG TCC ACA CAA TCT ACC CTT	624
193	P V L L P D N H Y L S T Q S T L	208
625	TTC AAA GAT CCC AAC GAA AAG AGA GAT CAC AUG ATC TAT TTT GGG TTT	672
209	S K D P N E K R D H M I Y F G F	224
673	GTA ACA GCT GCT GGG AAT ACA CHT GGC AUG GAT GAA TTA TAC AAA TAA	720
225	V T A A A I T H G M D E L Y K *	240

FIG. 6

1	ATG AGT AAA GGA GCA GAA CTT TTC ACT GGA	ATT GTC CCA ATT CTT ATT	48
1	M S K G A E L F T G	V P I L I	16
49	GAA TTA AAT GGT GAT GTT AAT GGG CAC AAA TTC TCT GTC AGT GGA GAG		96
17	E L N G D V N G H K F S V S G E		32
97	GGC GAA GGT GAT GCG ACA TAC GGA AAG TTA ACC CTT AAA TTT ATT TGC		144
33	G E G D A T Y G K L T L K F I C		48
145	ACT ACA GGA AAA CTA OCT GTT CCA TGG CCA ACA CTT GTC ACT ACT	GTC	192
49	T T G K L P V P W P T L V T T		64
193	TCT TAT GGT GTT CAA TGC TTT TCA AGA TAT CCA GAT CAT ATG AAA CAG		240
65	S Y G V Q C F S R Y P D H M K Q		80
241	CAT GAC TTC TTC AAG AGT GCC ATG OCT GAA GGT TAT ATA CAG GAA AGA		288
81	H D F F K S A M P E G Y I Q E R		96
289	ACT ATA TTT TTC	GAA GAT GAC GGG AAC TAC AAG TCG CGT GCT GAA GTC	336
97	T I F F	D D G N Y K S R A E V	112
337	AAG TTC GAA GGT GAT ACC CTG GTT AAT AGA ATT GAG TTA ACA GGT ACT		384
113	K F E G D T L V N R I E L T G T		128
385	GAT TTT AAA GAA GAT GGA AAC ATC CTT GGA AAT AAA ATG GAA TAC AAC		432
129	D F K E D G N I L G N K M E Y N		144
433	TAT AAC GCA CAT AAT GTA TAC ATC ATG ACA GAC AAA GCA AAA AAT GGA		480
145	Y N A H N V Y I M T D K A K N G		160
481	ATC AAA GGT AAC TTC AAA ATT AGA CAC AAC ATT GAA GAT GGA AGC GTT		528
161	I K V N F K I R H N I E D G S V		176
529	CAA CTT GCA GAC CAT TAT CAA CAA AAT ACT CCA ATT GGC GAT GGC OCT		576
177	Q L A D H Y Q Q N T P I G D G P		192
577	GTC CTT TTA CCA GAT AAC CAT TAC CTG TCC ACA CAA TCT ACC CIT TCC		624
193	V L L P D N H Y L S T Q S T L S		208
625	AAA GAT CCC AAC GAA AAG AGA GAT CAC ATG ATC TAT TTT	GGG TTT GGA	672
209	K D P N E K R D H M I Y F		224
673	ACA GCT GCT GCG ATT ACA CAT GGC ATG GAT GAA TTA TAC AAA TAA		717
225	T A A A I T H G M D E L Y K *		239

FIG. 7

1	ATG AGT AAA GGA GCA GAA CTT TTC ACT GGA ATT GTC CCA ATT CTT ATT	48
1	M S K G A E L F T G V P I L I	16
49	GAA TTA AAT GGT GAT GGT AAT GGG CAC AAA TTC TCT GTC AGT GGA GAG	96
17	E L N G D V N G H K F S V S G E	32
97	GGC GAA GGT GAT GCG ACA TAC GGA AAG TTA ACC CTT AAA TTT ATT TGC	144
33	G E G D A T Y G K L T L K F I C	48
145	ACT ACA GGA AAA CTA OCT GGT CCA TGG CCA ACA CTT GTC ACT ACT CTC	192
49	T T G K L P V P W P T L V T T	64
193	TCT TAT GGT GGT CAA TGC TTT TCA ACA TAT CCA GAT CAT ATG AAA CAG	240
65	S Y G V Q C F S R Y P D H M K Q	80
241	CAT GAC TTC TTC AAG AGT GGC ATG OCT GAA GGT TAT ATA CAG GAA AGA	288
81	H D F F K S A M P E G Y I Q E R	96
289	ACT ATA TTT TTC GAA GAT GAC GGG AAC TAC AAG TCG CGT GCT GAA GTC	336
97	T I F F D D G N Y K S R A E V	112
337	AAG TTC GAG GGT GAT ACC CIG GGT AAT ACA ATC GAG TTA ACA GGT ACT	384
113	K F E G D T L V N R I E L T G T	128
385	GAT TTT AAA GAA GAT GGA AAC ATC CTT GGA AAT AAA ATG GAA TAC AAC	432
129	D F K E D G N I L G N K M E Y N	144
433	TAT AAC GCA CAT AAT GTA TAC ATC ATG ACA GAC AAA GCA AAA AAT GCA	480
145	Y N A H N V Y I M T D K A K N G	160
481	ATC AAA GGT AAC TTC AAA ATT AGA CAC AAC ATT GAA GAT GGA AGC GGT	528
161	I K V N F K I R H N I E D G S V	176
529	CAA CTT GCA GAC CAT TAT CAA CAA AAT ACT CCA ATT GGC GAT GGC OCT	576
177	Q L A D H Y Q Q N T P I G D G P	192
577	GTC CTT TTA CCA GAT AAC CAT TAC CIG TCC ACA CAA TCT GGC CTT TCC	624
193	V L L P D N H Y L S T Q S L S	208
625	AAA GAT CCC AAC GAA AAG AGA GAT CAC ATG ATC TAT TTT GGG TTT GTA	672
209	K D P N E K R D H M I Y F F V	224
673	ACA GCT GCT GCG ATT ACA CAT GGC ATG GAT GAA CTA TAC AAA TAA	717
225	T A A A I T H G M D E L Y K *	239

FIG. 9

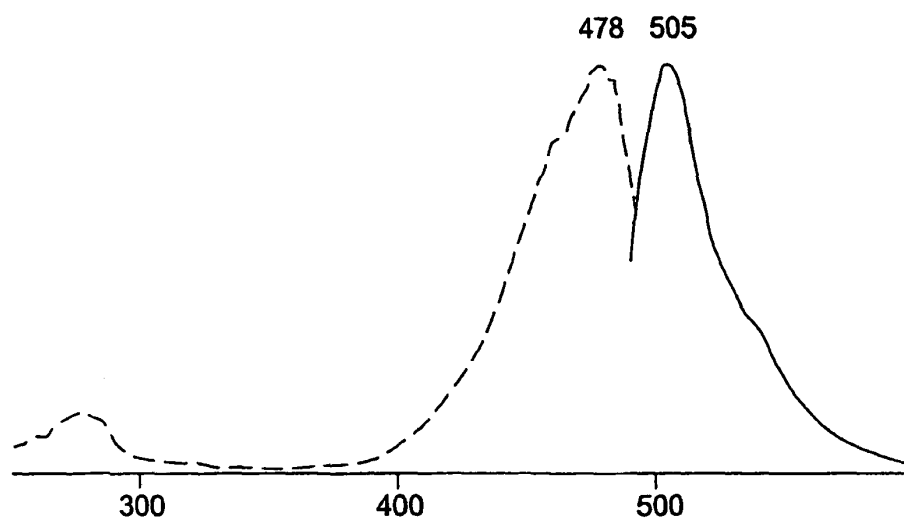


FIG. 10

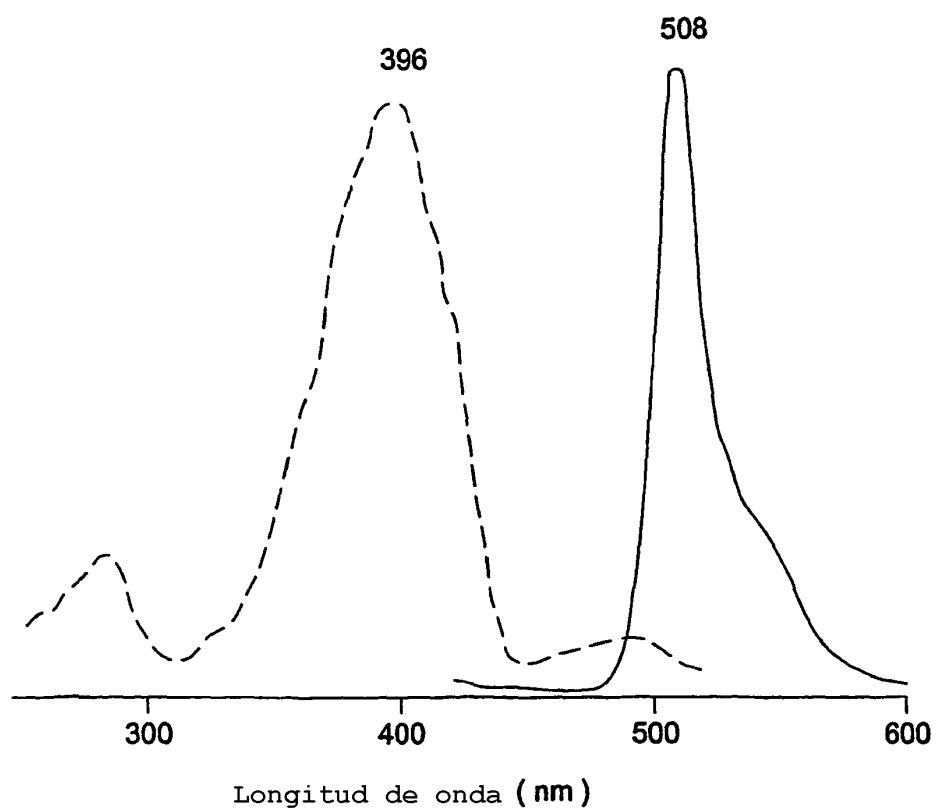


FIG. 16

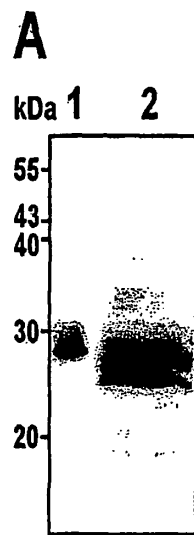


FIG. 11a

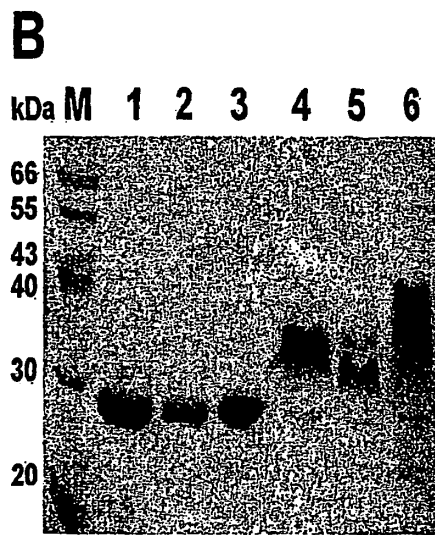


FIG. 11b



FIG. 11c

ES 2 340 154 T3

1	ATG AGT AAA GGA GCA GAA CTT TTC ACT GGA	48
1	M S K G A E L F T G	16
49	GAA TTA AAT GGT GAT GTT AAT GGG CAC AAA TTC TCT GTC AGT GGA GAG	96
17	E L N G D V N G H K F S V S G E	32
97	GGC GAA GGT GAT GCG ACA TAC GGA AAG TTA ACC CTT AAA TTT ATT TGC	144
33	G E G D A T Y G K L T L K F I C	48
145	ACT ACA GGA AAA CTA OCT GTT CCA TGG CCA ACA CTT GTC ACT ACT	192
49	T T G K L P V P W P T L V T T	64
193	TCT TAT GGT GTT CAA TGC TTT TCA AGA TAT CCA GAT CAT ATG AAA CAG	240
65	S Y G V Q C F S R Y P D H M K Q	80
241	CAT GAC TTC TTC AAG AGT GCC ATG OCT GAA GGT TAT ATA CAG GAA ACA	288
81	H D F F K S A M P E G Y I Q E R	96
289	ACT ATA TTT TTC GAA GAT GAC GGG AAC TAC AAG TCG CGT GCT GAA GTC	336
97	T I F F E D D G N Y K S R A E V	112
337	AAG TTC GAG GGT GAT ACC CTG GTT AAT AGA ATC GAG TTA ACA GGT ACT	384
113	K F E G D T L V N R I E L T G T	128
385	GAT TTT AAA GAA GAT GGA AAC ATC CTT GGA AAT AAA ATG GAA TAC AAC	432
129	D F K E D G N I L G N K M E Y N	144
433	TAT AAC GCA CAT AAT GTA TAC ATC ATG ACA GAC AAA GCA AAA AAT GGA	480
145	Y N A H N V Y I M T D K A K N G	160
481	ATC AAA GTT AAC TTC AAA ATT AGA CAC AAC ATT GAA GAT GGA AGC GTT	528
161	I K V N F K I R H N I E D G S V	176
529	CAA CTT GCA GAC CAT TAT CAA CAA AAT ACT CCA ATT GGC GAT GGC CCT	576
177	Q L A D H Y Q Q N T P I G D G P	192
577	GTC CTT TTA CCA GAT AAC CAT TAC CTG TOC ACA CAA TCT	624
193	V L L P D N H Y L S T Q S	208
625	AAA GAT CCC AAC GAA AAG AGA GAT CAC ATG ATC TAT TTT GAG TTT GTA	672
209	K D P N E K R D H M I Y F E F V	224
673	ACA GCT GCT GCG ATT ACA CAT GGC ATG GAT GAA CTA TAC AAA TAA	717
225	T A A A I T H G M D E L Y K *	239

FIG. 12

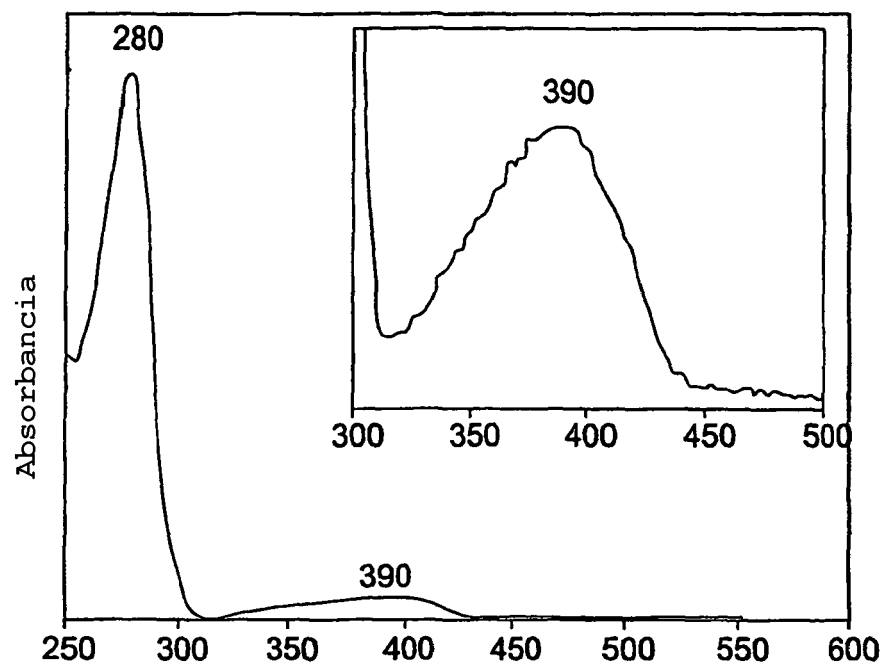


FIG. 13a

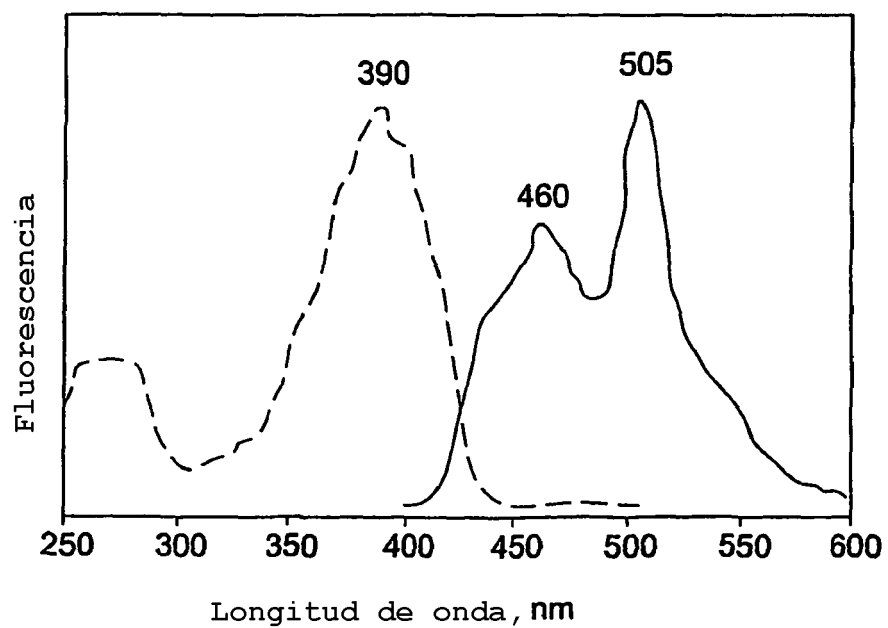


FIG. 13b

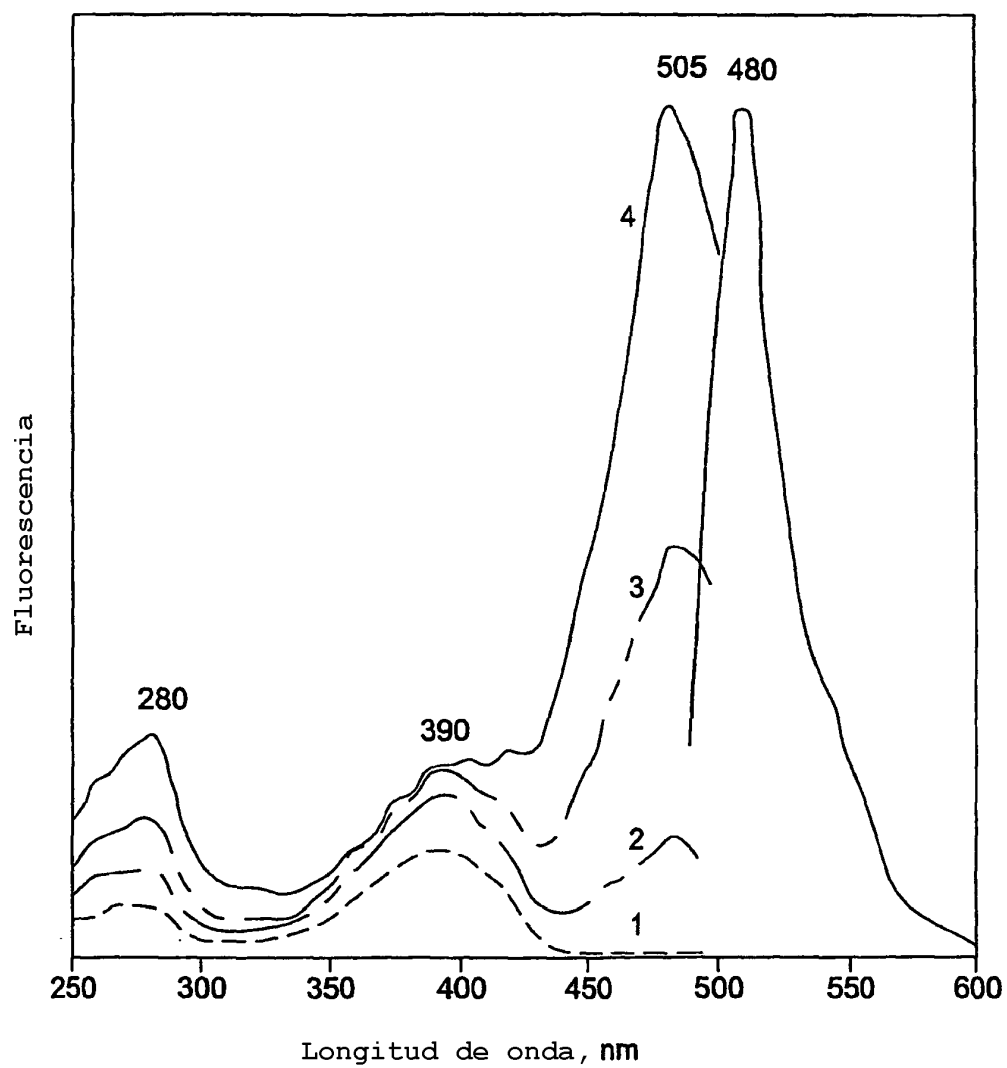


FIG. 14

1	ATG AGT AAA GGA GCA GAA CTT TTC ACT GGA ATT GTC CCA ATT CTT ATT	48
1	M S K G A E L F T G V P I L I	16
49	GAA TTA AAT GGT GAT GGT AAT GGG CAC AAA TTC TCT GTC AGT GGA GAG	96
17	E L N G D V N G H K F S V S G E	32
97	GGC GAA GGT GAT GGG ACA TAC GGA AAG TTA ACC CTT AAA TTT ATT TGC	144
33	G E G D A T Y G K L T L K F I C	48
145	ACT ACA GGA AAA CTA CCT GGT CCA TGG CCA ACA CTT GTC ACT ACT CTC	192
49	T T G K L P V P W P T L V T T	64
193	TCT TAT GGT GGT CAA TGC TTT TCA AGA TAT CCA GAT CAT ATG AAA CAG	240
65	S Y G V Q C F S R Y P D H M K Q	80
241	CAT GAC TTC TTC AAG AGT GGC ATG CCT GAA GGT TAT ATA CAG GAA AGA	288
81	H D F F K S A M P E G Y I Q E R	96
289	ACT ATA TTT TTC GAA GAT GAC GGG AAC TAC AAG TCG CGT GCT GAA GTC	336
97	T I F F E D D G N Y K S R A E V	112
337	AAG TTC GAG GGT GAT ACC CAG GGT AAT AGA ATC GAG TTA ACA GGT ACT	384
113	K F E G D T L V N R I E L T G T	128
385	GAT TTT AAA GAA GAT GGA AAC ATC CTT GGA AAT AAA ATG GAA TAC AAC	432
129	D F K E D G N I L G N K M E Y N	144
433	TAT AAC GCA CAT AAT GTA TAC ATC ATG ACA GAC AAA GCA AAA AAT GGA	480
145	Y N A H N V Y I M T D K A K N G	160
481	ATC AAA GGT AAC TTC AAA ATT AGA CAC AAC ATT GAA GAT GGA AGC GGT	528
161	I K V N F K I R H N I E D G S V	176
529	CAA CTT GCA GAC CAT TAT CAA CAA AAT ACT CCA ATT GGC GAT GGC CCT	576
177	Q L A D H Y Q Q N T P I G D G P	192
577	GTC CTT TTA CCA GAT AAC CAT TAC CAG TCC ACA CAA TCT GGC CTT TCC	624
193	V L L P D N H Y L S T Q S A L S	208
625	AAA GAT CCC AAC GAA AAG AGA GAT CAC ATG ATC CAG TTT GAG TTT GTA	672
209	K D P N E K R D H M I F E F V	224
673	ACA GCT GCT GCG ATT ACA CAT GGC ATG GAT GAA CTA TAC AAA TAA	717
225	T A A A I T H G M D E L Y K *	239

FIG. 15

1	ATG AGT AAA GGA GCA GAA CTT TTC ACT GGA ATT GTC OCA ATT CTT ATT	48
1	M S K G A E L F T G I V P I L I	16
49	GAA TTA AAT GGT GAT GTT AAT GGG CAC AAA TTC TCT GTC AGT GGA GAG	96
17	E L N G D V N G H K F S V S G E	32
97	GGC GAA GGT GAT GGG ACA TAC GGA AAG TTA ACC CTT AAA TTT ATT TGC	144
33	G E G D A T Y G K L T L K F I C	48
145	ACT ACA GGA AAA CTA CCT GTT OCA TGG OCA ACA CTT GTC ACT ACT GTC	192
49	T T G K L P V P W P T L V T T E	64
193	TCT TAT GGT GGT CAA TGC TTT TCA AGA TAT OCA GAT CAT ATG AAA CAG	240
65	S Y G A Q C F S R Y P D H M K Q	80
241	CAT CAC TTC TTC AAG AGT GGC ATG OCT GAA GGT TAT ATA CAG GAA AGA	288
81	H D F F K S A M P E G Y I Q E R	96
289	ACT ATA TTT TTC GAA GAT GAC GGG AAC TAC AAG TCG CGT GCT GAA GTC	336
97	T I F F E D D G N Y K S R A E V	112
337	AAG TTC AAG GGT GAT ACC CTG GTT AAT AGA ATC CAG TTA ACA GGT ACT	384
113	K F K G D T L V N R I E L T G T	128
385	GAT TTT AAA GAA GAT GGA AAC ATC CTT GGA AAT AAA ATG GAA TAC AAC	432
129	D F K E D G N I L G N K M E Y N	144
433	TAT AAC GCA CAG AAT GTA TAC ATC ATG ACA GAC AAA GCA AAA AAT GGA	480
145	Y N A Q N V Y I M T D K A K N G	160
481	ATC AAA GIT AAC TTC AAA ATT AGA CAC AAC ATT GAA GAT GGA AGC GIT	528
161	I K V N F K I R H N I E D G S V	176
529	CAA CTT GCA GAC CAT TAT CAA CAA AAT ACT OCA ATT GGC GAT GGC OCT	576
177	Q L A D H Y Q Q N T P I G D G P	192
577	GTC CTT TTA OCA GAT AAC CAT TAC CTG TOC ACA CAA TCT GGC CTT TOC	624
193	V L L P D N H Y L S T Q S E L S	208
625	AAA GAT CCC AAC GAA AAG AGA GAT CAC ATG ATC CAG CAG GAG TTT GTA	672
209	K D P N E K R D H M I L L E F V	224
673	ACA GCT GCT GCG ATT ACA CAT GGC ATG GAT GAA CTA TAC GAA TAA	717
225	T A A A I T H G M D E L Y Q *	239

FIG. 17

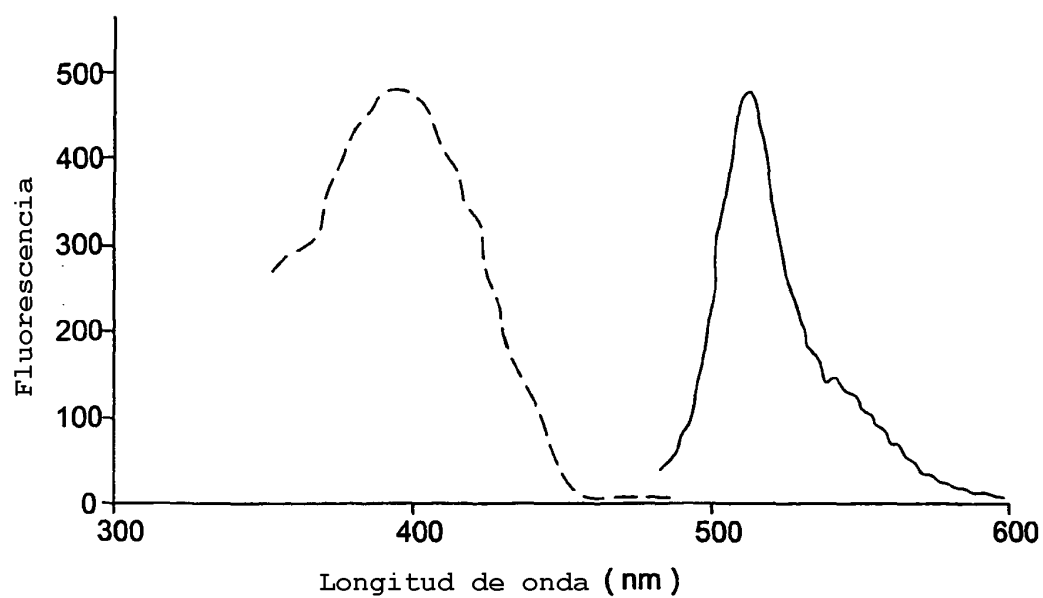


FIG. 18a

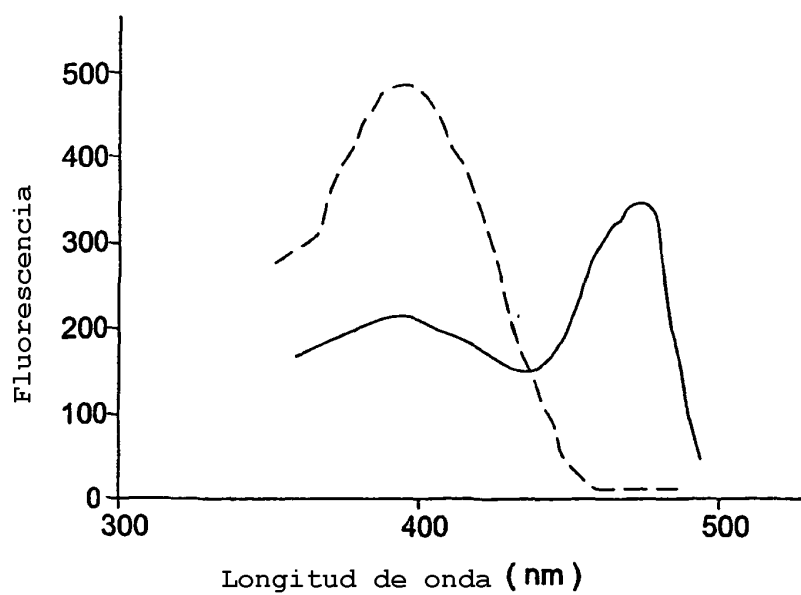


FIG. 18b

1	ATG AGT AAA GGA GCA GAA CTT TTC ACT GGA ATT GTC CCA ATT CTT ATT	48
1	M S K G A E L F T G V P I L I	16
49	GAA TTA AAT GGT GAT GTT AAT GGG CAC AAA TTC TCT GTC AGT GGA GAG	96
17	E L N G D V N G H K F S V S G E	32
97	GGC GAA GGT GAT GGG ACA TAC GGA AAG TTA ACC CTT AAA TTT ATT TGC	144
33	G E G D A T Y G K L T L K F I C	48
145	ACT ACA GGA AAA CTA CCT GTT CCA TGG CCA ACA CTT GTC GCT ACT GTC	192
49	T T G K L P V P W P T L V A T I	64
193	TCT TAT GGT GTT CAA TGC TTT TCA AGA TAT CCA GAT CAT ATG AAA CAG	240
65	S Y G V Q C F S R Y P D H M K Q	80
241	CAT GAC TTC TTC AAG AGT GGC ATG CCT GAA GGT TAT ATA CAG GAA AGA	288
81	H D F F K S A M P E G Y I Q E R	96
289	ACT ATA TTT TTC TAA GAT GAC GGG AAC TAC AAG TCG CGT GCT GAA GTC	336
97	T I F F E D D G N Y K S R A E V	112
337	AAG TTC GAG GGT GAT ACC CTG GTT AGT AGA ATC GAG TTA ACA GGT ACT	384
113	K F E G D T L V S R I E L T G T	128
385	GAT TTT AAA GAA CAT GGA AAC ATC CTT GGA AAT AAA ATG GAA TAC AAC	432
129	D F K E D G N I L G N K M E Y N	144
433	TAT AAC GCA TAT AAT GTA TAC ATC ATG ACA GAC AAA GCA AAA AAT GGA	480
145	Y N A E N V Y I M T D K A K N G	160
481	ATC AAA GAT AAC TTC AAA ATT AGA CAC AAC ATT TAA CAT GGA AGC GTT	528
161	I K V N F K I R H N I K D G S V	176
529	CAA CTT GCA GAC CAT TAT CAA CAA AAT ACT CCA ATT GGC GAT GGC CCT	576
177	Q L A D H Y Q Q N T P I G D G P	192
577	GTC CTT TTA CCA GAT AAC CAT TAC CIG TCC ACA CAA TCT GGC CTT TCC	624
193	V L L P D N H Y L S T Q S A L S	208
625	AAA GAT CCT AAC GAA AAG AGA GAT CAC ATG ATC TAT TTT GAG TTT GTA	672
209	K D P N E K R D H M I Y F E F V	224
673	ACA GCT GCT GGG ATT ACA CAT GGC ATG GAT GAA CTA TAC AAA TAA	717
225	T A A A I T H G M D E L Y K *	239

FIG. 19

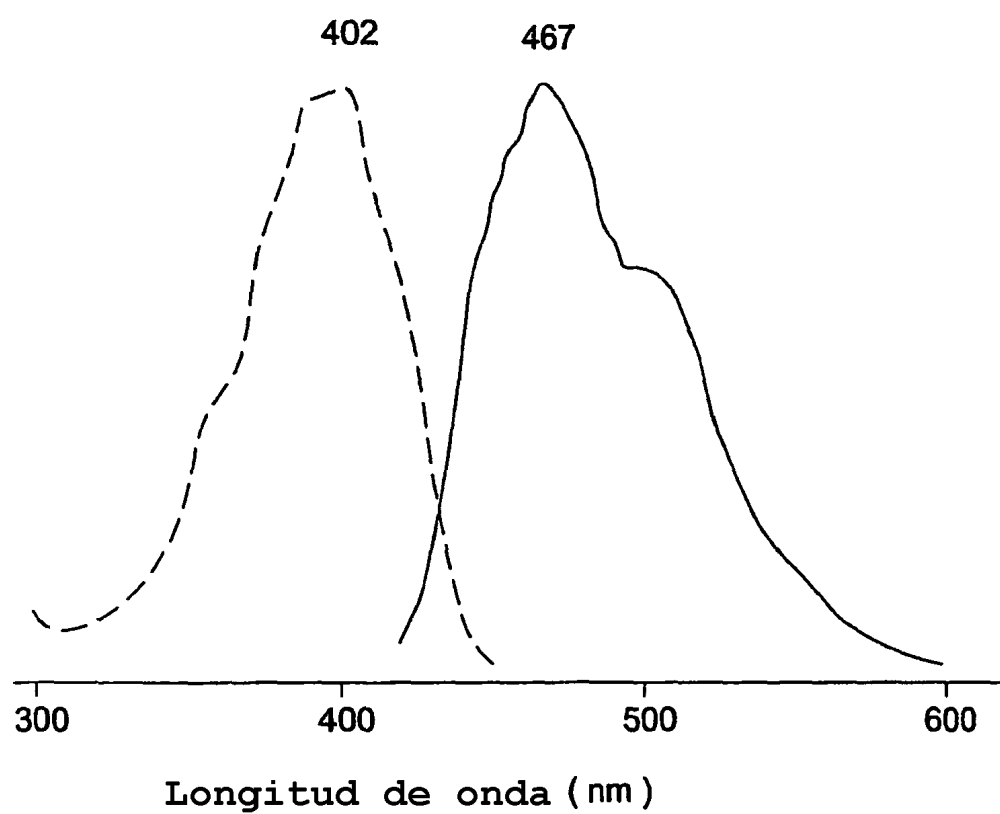


FIG. 20

1	ATG AGT AAA GGA GCA GAA CTT TTC ACT GGA ATT GTC CCA ATT CTT ATT	48
1	M S K G A E L F T G V P I L I	16
49	GAA TTA AAT GGT GAT GTT AAT GGG CAC AAA TTC TCT GTC AGT GGA GAG	96
17	E L N G D V N G H K F S V S G E	32
97	GCC GAA GGT GAT GCG ACA TAC GGA AAG TTA ACC CTT AAA TTT ATT TGC	144
33	G E G D A T Y G K L T L K F I C	48
145	ACT ACA GGA AAA CTA CCT GTT CCA TGG CCA ACA CTT GTC ACT ACT GTC	192
49	T T G K L P V P W P T L V T T	64
193	TCT TAT GGT GTT CAA TGC TTT TCA AGA TAT CCA GAT CAT ATG AAA CAG	240
65	S Y G V Q C F S R Y P D H M K Q	80
241	CAT GAC TTC TTC AAG AGT GCC ATG CCT GAA GGT TAT ATA CAG GAA AGA	288
81	H D F F K S A M P E G Y I Q E R	96
289	ACT ATA TTT TTC GAA GAT GAC GGG AAC TAC AAG TCG CGT GCT GAA GTC	336
97	T I F F D D G N Y K S R A E V	112
337	AAG TTC GAG GGT GAT ACC CAG GTT AAT AGA ATC GAG TTA ACA GGT ACT	384
113	K F E G D T L V N R I E L T G T	128
385	GAT TTT AAA GAA GAT GGA AAC ATC CTT GCA AAT AAA ATG GAA TAC AAC	432
129	D F K E D G N I L G N K M E Y N	144
433	TAT AAC GCA TGT AAT GTA TAC ATC ATG ACA GAC AAA GCA AAA AAT GGA	480
145	Y N A S N V Y I M T D K A K N G	160
481	ATC AAA GTT AAC TTG AAA ATT AGA CAC AAC ATT GCA GAT GGA AGC GTT	528
161	I K V N K I R H N I A D G S V	176
529	CAA CTT GCA GAC CAT TAT CAA CAA AAT ACT CCA ATT GGC GAT GGC CCT	576
177	Q L A D H Y Q Q N T P I G D G P	192
577	GTC CTT TTA CCA GAT AAC CAT TAC CAG TCC ACA CAA TCT GGC CTT TCC	624
193	V L L P D N H Y L S T Q S A L S	208
625	AAA GAT CCC AAC GAA AAG AGA GAT CAC ATG ATC TAT TTT GAG TTT GTA	672
209	K D P N E K R D H M I Y F E F V	224
673	ACA GCT GCT GCG ATT ACA CAT GGC ATG GAT GAA CTA TAC AAA TAA	717
225	T A A A I T H G M D E L I K *	239

FIG. 21

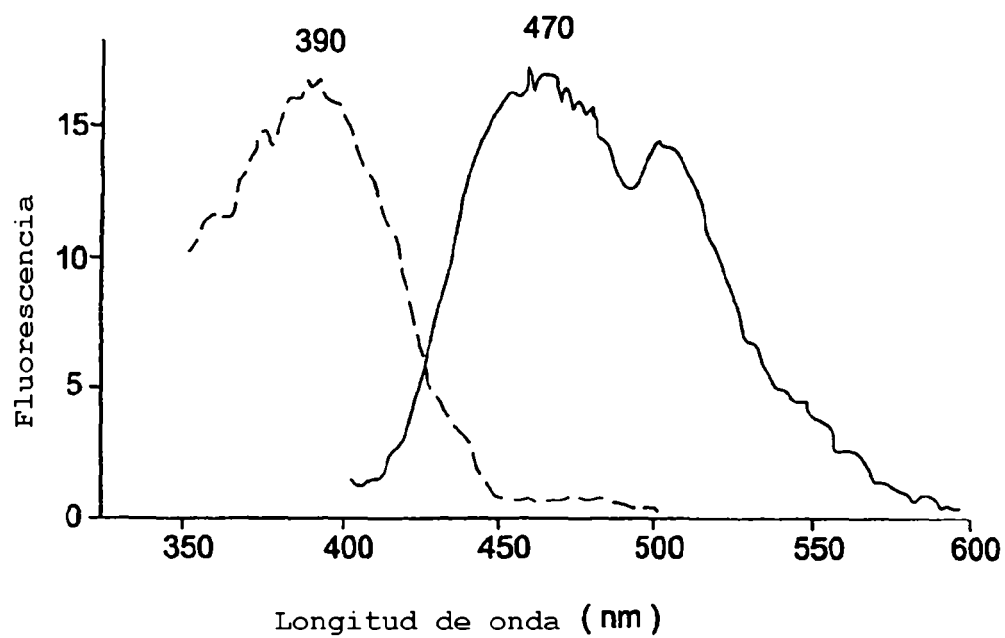


FIG. 22a

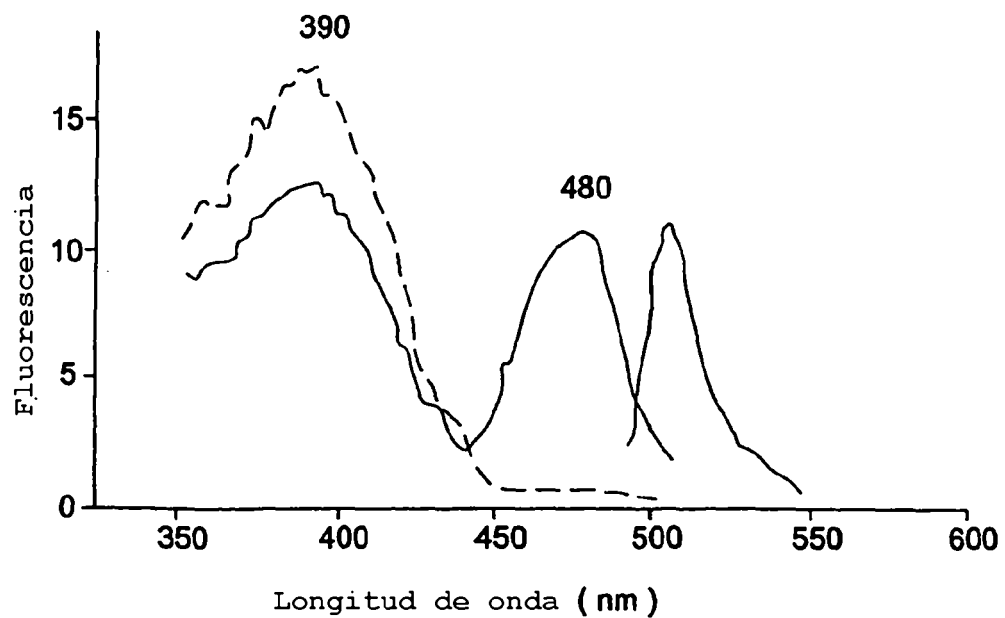


FIG. 22b

1	ATGAGCAAGGGCGCCGAGCTGTTCAACGGCATCGTGCCCATCCTGATC	48
1	M S K G A E L F T G I V P I L I	16
49	GAGCTGAATGGCGATGTGAATGGCCACAAGTTCAGCGTGAGCGGCGAG	96
17	E L N G D V N G H K F S V S G E	32
97	GGCGAGGGCGATGCCACCTACGGCAAGCTGACCCCTGAAGTTCATCTGC	144
33	G E G D A T Y G K L T L K F I C	48
145	ACCACCGGCAAGCTGCCTGTGCCCTGGCCCAACCTGGTGACCACCCCTG	192
49	T T G K L P V P W P T L V T T L	64
193	AGCTACGGCGTGCAAGTGCTTCTCACGCTACCCCGATCACATGAAGCAG	240
65	S Y G V Q C F S R Y P D H M K Q	80
241	CACGACTTCCTTCAAGAGCGCCATGCCTGAGGGCTACATCCAGGAGCGC	288
81	H D F F K S A M P E G Y I Q E R	96
289	ACCATCTTCTTCGAGGATGACGGCAACTACAAGTCGCGCGCCGAGGTG	336
97	T I F F E D D G N Y K S R A E V	112
337	AAGTTCGAGGGCGATACCCCTGGTGAATCGCATCGAGCTGACCGGCACC	384
113	K F E G D T L V N R I E L T G T	128
385	GATTTCAAGGAGGATGGCAACATCCTGGGCAATAAGATGGAGTACAAC	432
129	D F K E D G N I L G N K M E Y N	144
433	TACAACGCCCCACAATGTGTACATCATGACCGACAAGGCCAAGAATGGC	480
145	Y N A H N V Y I M T D K A K N G	160
481	ATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGATGGCAGCGTG	528
161	I K V N F K I R H N I E D G S V	176
529	CAGCTGGCCGACCACTACCAGCAGAATACCCCCATCGGCGATGGCCCT	576
177	Q L A D H Y Q Q N T P I G D G P	192
577	GTGCTGCTGCCCGATAACCACTACCTGTCCACCCAGAGCGCCCTGTCC	624
193	V L L P D N H Y L S T Q S A L S	208
625	AAGGACCCCCAACGAGAAGCGCGATCACATGATCTACTTCGGCTTCGTG	672
209	K D P N E K R D H M I Y F G F V	224
673	ACCGCCGCGCCATCACCCACGGCATGGATGAGCTGTACAAGTGA	717
225	T A A A I T H G M D E L Y K *	239

FIG. 23

FIG. 24a

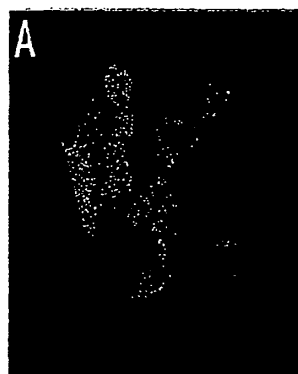


FIG. 24b



FIG. 24c



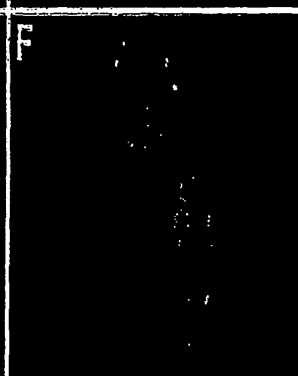
FIG. 24d



FIG. 24e



FIG. 24f



ES 2 340 154 T3

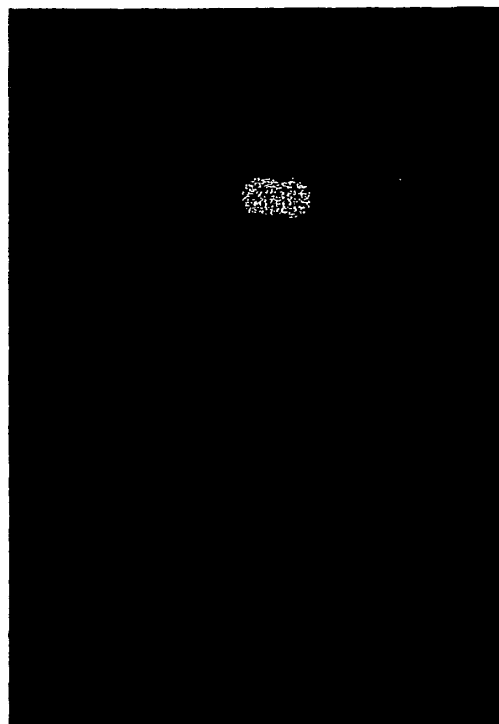


FIG. 25